University of Nottingham
UK | CHINA | MALAYSIA

Disentangling diphtheria: An investigation into its persistence worldwide

Juniorcaius Ikejezie, BSc, MSc

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

December 2024

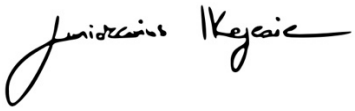
School of Medicine

University of Nottingham

Table of contents

Declaration

I, hereby, declare that this thesis is my own work. It is based on original research, which I conducted under the supervision of Prof. Sarah Lewis, Dr. Tessa Langley, Dr. Revati Phalkey, and Dr. Donal Bisanzio. This thesis has not been submitted elsewhere for any degree, diploma or other qualification. All the authors and works to which reference has been made are fully acknowledged.



Juniorcaius Ikejezie

December 2024

Abstract

Introduction: Diphtheria is a contagious disease primarily caused by *Corynebacterium diphtheriae*, which typically affects the respiratory tract and less often the skin. Transmission occurs via exposure to secretions of an infected person. Generally, respiratory diphtheria has a case fatality rate of 5–10%, with higher rates seen among certain groups (e.g., untreated, unvaccinated individuals). After years of decline, diphtheria is experiencing a resurgence, with 24,778 cases reported in 2023 – the highest global case count in over 20 years. The disease has severely affected Haiti, a country that is also facing vulnerability to natural disasters, concurrent health crises, and civil unrest. Despite its impact, diphtheria is still relatively understudied, with questions remaining about the determinants of the observed epidemiological trends and patterns. This PhD thesis aimed to explore drivers of global diphtheria persistence to fill current knowledge gaps.

Methods: The thesis consisted of three studies. The first study was a systematic review and meta-analysis to identify risk factors for diphtheria. The second study was a spatial analysis of the epidemiology of the disease in Haiti between December 2014 and June 2021 to examine the spatiotemporal distribution of cases, detect diphtheria hotspots, and identify potential risk factors for the disease. The third study was a mixed-methods analysis of posts shared from January 2012 to December 2022 on X (formerly Twitter) to understand global perceptions and attitudes towards diphtheria.

Results: The review identified three potential risk factors for diphtheria: incomplete vaccination, contact with a person with skin lesions, and low

diphtheria knowledge. Evidence for other risk factors was less conclusive. The spatial analysis revealed that the reported diphtheria case rate exhibited spatial variability in Haiti, with nine out of 140 communes being classified as hotspots. This rate was positively associated with the number of healthcare facilities per 100,000 inhabitants and the proportion of urban population, while it had a negative association with female literacy. The mixed-methods analysis showed that a diversity of voices participates in diphtheria-related conversations on X, leading to the identification of various themes – with vaccination being the predominant topic of interest. The analysis also highlighted posts spreading false information alongside differences in tone and content between high-burden and low-burden countries, mirroring variations in healthcare priorities and challenges.

Conclusion: This thesis illustrates the complex nature of diphtheria, as evidenced by the multitude of biological, immunological, environmental, socioeconomic, behavioural, and informational factors that potentially contribute to the persistence of the disease. The in-depth investigation of diphtheria in Haiti highlighted the compounded difficulties in controlling the disease during a humanitarian crisis. Findings from this case study, alongside the risk factors and themes identified through the systematic review and mixed-methods analysis, can inform future research and interventions not only in Haiti but also in other countries facing similar challenges. While remaining context-specific, these interventions should seek to raise vaccination coverage, scale up access to healthcare services, strengthen surveillance systems, increase health literacy, and improve sanitation and hygiene conditions. Innovative tools like spatial analysis and social media monitoring can enhance surveillance, by facilitating the detection of disease trends, at-risk populations, and emerging public concerns.

Acknowledgements

Undertaking this PhD has been one of the most demanding yet fulfilling experiences of my life. The completion of this thesis would simply have been impossible without the contributions of many exceptional people.

First and foremost, I would like to wholeheartedly thank my supervisors, Prof. Sarah Lewis, Dr. Tessa Langley, Dr. Revati Phalkey, and Dr. Donal Bisanzio, for their expert guidance, insightful feedback, and steadfast mentorship. I am incredibly grateful for having had the opportunity to conduct research under their supervision.

I am sincerely thankful to the staff of the University of Nottingham's School of Medicine for providing the resources and tools necessary to perform rigorous research. Particular appreciation also goes to my fellow students for the camaraderie, stimulating discussions, and words of encouragement, which helped me persevere during difficult times. I am especially grateful to Drs. Busola Adebusoye and Winifred Ekezie for their invaluable contributions to the systematic review that underpins this thesis.

I am profoundly indebted to Haiti's Ministry of Public Health and Population, as well as researchers at Boston Children's Hospital, whose data were key to the realization of two of the three studies that form this research project. Their commitment to improving public health through data sharing and knowledge dissemination has been truly uplifting.

A special mention goes to the healthcare professionals in Haiti and around the world who are on the frontline of the fight against diphtheria. Their devotion to saving lives motivates me and serves as a constant reminder of the fundamental role of public health.

Lastly, I dedicate this thesis to my family. To my parents, Dr. Kaius Ikejezie and Amaka Agatha Akabueze, for their constant wisdom, sacrifices, and love; to my brother, Carl Ikejezie, and his girlfriend, Nadia Afework, for their continual advice, humour, and love; and to my partner, Mikal Yonas, for her endless kindness, affection, and love.

Training courses and events

Year	Course / event	Credit points
1	Advanced Statistical Methods	14
	Data Organisation and Management in Epidemiology (DOME)	
	Health Economics	
	Principles and Process of Comprehensive Systematic Review	
	Systematic Review Masterclass	
	Research Methods in Epidemiology with Basic Statistics	13
	Introduction to R Training Workshop	6
	GIS Workshop	2
	Introduction to Statistics with SPSS	2
	Medicine and Health Sciences Faculty Postgraduate Research Forum	2
	Critical Appraisal of Scientific Literature 1 (Non-clinical)	1
	Introduction to Endnote for Researchers	1
	Researcher Information Skills for Medicine & Health Sciences	1
	UK Data Service Event	1
	Preparing for Your Confirmation Review	(Attended, no credits awarded)
	Writing an Impact Statement (Medicine and Health Sciences)	(Attended, no credits awarded)
2	Sue Watson Oral Presentation Event	1
3	Preparing for your Viva (Webinar)	1
Total		**45**

Conference presentations

IKEJEZIE, J., PHALKEY, R, LEWIS, S., LANGLEY, T., AND BISANZIO, D. Disentangling diphtheria: An investigation into its persistence in Haiti. Medicine and Health Sciences Faculty Postgraduate Research Forum. University of Nottingham. 25 June 2019.

IKEJEZIE, J., LANGLEY, T., LEWIS, S., BISANZIO, D., AND PHALKEY, R. The epidemiology of diphtheria in Haiti, 2014 – 2021: A spatial modeling analysis. Faculty Research Forum. University of Nottingham. 7 December 2021. Awarded first prize: £75.

List of publications and contributions

IKEJEZIE, J., LANGLEY, T., LEWIS, S., BISANZIO, D., AND PHALKEY, R. (2022). The epidemiology of diphtheria in Haiti, December 2014 – June 2021: a spatial modeling analysis. *PLoS ONE* 17(8): e0273398. https://doi.org/10.1371/journal.pone.0273398.

- **Juniorcaius Ikejezie**: Designed the study protocol; obtained and managed epidemiological data from Haiti's Ministry of Public Health and Population; collected additional data from other sources; performed data analysis; interpreted results; wrote and revised the manuscript.
- **Tessa Langley, Sarah Lewis, Donal Bisanzio, Revati Phalkey:** Provided guidance on the study design and methodology; reviewed the manuscript and offered critical feedback.

IKEJEZIE, J., ADEBUSOYE, B., EKEZIE, W., LANGLEY, T., LEWIS, S., AND PHALKEY, R. (2023). Modifiable risk factors for diphtheria: a systematic review and meta-analysis. *Global Epidemiology*, 100100.

- **Juniorcaius Ikejezie**: Designed the study protocol; conducted the literature search; screened all titles, abstracts, and full texts; extracted data from included studies; evaluated the methodological quality of the studies; performed data analysis; interpreted results; wrote and revised the manuscript.
- **Busola Adebusoye:** Screened abstracts and full texts; extracted data from included studies; reviewed the manuscript and offered critical feedback.

- **Winifred Ekezie:** Evaluated the methodological quality of the studies; reviewed the manuscript and offered critical feedback.
- **Tessa Langley, Sarah Lewis, Revati Phalkey:** Provided guidance on the study design, methodology, and interpretation of findings; reviewed the manuscript and offered critical feedback.

IKEJEZIE, J., LEWIS, S., LANGLEY, T., TULI, G., SEWALK, K., REMMEL, C., BROWNSTEIN, J. S., BISANZIO, D. (2024). The public discourse on diphtheria: A quantitative and qualitative analysis of 2012–2022 data from X (formerly Twitter). Under review at *Perspectives in Public Health*.

- **Juniorcaius Ikejezie**: Designed the study protocol; preprocessed social media data; conducted quantitative and qualitative analyses; interpreted results; wrote and revised the manuscript.
- **Tuli Gaurav, Kara Sewalk, Christopher Remmel, John Brownstein:** Collected social media data; reviewed the manuscript and offered critical feedback.
- **Tessa Langley, Sarah Lewis, Donal Bisanzio:** Provided guidance on the study design, methodology, and interpretation of findings; reviewed the manuscript and offered critical feedback.

List of figures

List of tables

List of abbreviations

AFR	African Region
AMR	Antimicrobial resistance
AMR	Region of the Americas
AIC	Akaike information criterion
AIC_c	Corrected Akaike information criterion
AIDS	Acquired Immunodeficiency Syndrome
BC	Before Christ
BCH	Boston Children's Hospital
CASP	Critical Appraisal Skills Programme
CFR	Case fatality rate
COVID-19	Coronavirus disease 2019
CS	Coherence score
DAT	Diphtheria antitoxin
DELR	Haiti's Directorate of Epidemiology, Laboratory and Research
DHS	Demographic and Health Surveys
DT	Diphtheria-tetanus vaccine
DTP	Diphtheria-tetanus-pertussis-containing vaccine
DTP3	Diphtheria-tetanus-pertussis-containing vaccine, third dose
DTP4	Diphtheria-tetanus-pertussis-containing vaccine, fourth dose
EEA	European Economic Area
EMR	Eastern Mediterranean Region

EPI	Expanded Programme on Immunization
EU	European Union
EUR	European Region
EVD	Ebola virus disease
EVM	Effective vaccination management
EQA	External quality assurance
GDP	Gross domestic product
GIS	Geographic information systems
GRADE	Grading of Recommendations Assessment, Development, and Evaluation
GWR	Geographically weighted regression
HCWs	Healthcare workers
HDX	Humanitarian Data Exchange
HICs	High-income countries
HPV	Human papillomavirus
HTG	Haitian gourde
IHSI	Haitian Institute of Statistics and Informatics
IIS	Immunization information systems
JBI	Joanna Briggs Institute
LDA	Latent Dirichlet Allocation
LICs	Low-income countries
LISA	Local indicators of spatial autocorrelation
LMICs	Low- and middle-income countries
MeSH	Medical Subject Headings
MSPP	Haiti's Ministry of Public Health and Population
N/A	Not available
NESN	National Epidemiologic Surveillance Network
NGOs	Nongovernmental organizations

NOS	Newcastle-Ottawa Scale
OLS	Ordinary least squares
OR	Odds ratio
PAHO	Pan American Health Organization
PCR	Polymerase chain reaction
PDR	People's Democratic Republic
PHIL	Public Health Image Library
POR	Pooled odds ratio
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROMESS	Program on Essential Medicine and Supplies
R^2	The coefficient of determination
R&D	Research and development
RCTs	Randomized controlled trials
REC	Research Ethics Committee
ROBINS-I	Risk of Bias in Non-randomized Studies - of Interventions
RR	Risk ratio
SAR	Spatial Autoregressive Model
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SEAR	South-East Asia Region
SEM	Spatial Error Model
SES	Socioeconomic status
TB	Tuberculosis
TM	Topic modeling
UN	United Nations
UN OCHA	United Nations Office for the Coordination of Humanitarian Affairs

UNICEF	United Nations International Children's Emergency Fund
UoN	University of Nottingham
US	United States
U.S. CDC	United States Centers for Disease Control and Prevention
VIF	Variance inflation factor
WHO	World Health Organization
WHOLIS	WHO Library Information System
WPR	Western Pacific Region

Chapter 1. Introduction

1.1. Infectious disease surveillance

1.1.1. Importance of infectious disease surveillance

The history of humanity has been inextricably linked with recurring catastrophic infectious disease outbreaks that have caused widespread mortality, while disrupting social structures, economies, and norms across the world [1, 2]. Historical examples of devastating epidemics and pandemics over the centuries are numerous and include the plague that ravaged Athens in 430 before Christ (BC), leading to the death of 75,000–100,000 people (25% of the city-state's population) [1, 3]; the medieval Black Death, which is presumed to have decimated approximately 30–50% of Europe's population from 1347 to 1351 [2, 4]; the 1918 influenza pandemic, also called the "Spanish flu," which resulted in the death of over 17 million people globally [1, 5]; and the recent pandemic of Coronavirus disease 2019 (COVID-19), which caused over 772 million cases and about seven million deaths between 2020 and 2023 [1, 6].

These examples are stark reminders of the critical importance of infectious disease surveillance. This is not simply a scholarly endeavor, but a vital undertaking for the protection of human lives, which helps to understand the complex processes of pathogen transmission, identify the determinants of disease spread, and formulate necessary prevention and containment interventions [7]. This pillar of public health is characterized by a multidisciplinary approach, integrating different scientific disciplines, such as biostatistics, epidemiology, immunology,

microbiology, and social sciences. Each of these disciplines provides distinct perspectives into the nature of infectious diseases [7]. As global interconnectedness increases, so do the opportunities for pathogens to spread worldwide. Therefore, infectious disease surveillance remains a pivotal element for the maintenance of international health security [7].

1.1.2. Traditional and innovative surveillance methods

Historically, infectious disease surveillance was predominantly empirical and relied heavily on the observations of clinical manifestations by physicians, who used techniques of different effectiveness [2, 8]. Examples of these techniques include visually inspecting a patient to detect signs of illness (e.g., skin discolorations, rashes, lesions), palpating the body to identify physical abnormalities (e.g., swellings, tumours), and listening to the internal sounds of the body (a method known as auscultation) to assess the functioning of organs like the heart, intestines, and lungs [2, 8].

These examinations used to be conducted without in-depth knowledge of the underlying pathogenic causes of diseases prior to the discovery of microorganisms by Antonie van Leeuwenhoek in the 17th century and the pioneering work in the 19th century of scientists like Ignaz Semmelweis, John Snow, Louis Pasteur, and Robert Koch [2, 8]. Collectively, their research contributed to the development and application of the germ theory of disease, which asserts that infectious diseases are caused by pathogenic organisms [2, 8]. This theory revolutionized the understanding of disease processes, paving the way for modern public health practices and providing the scientific basis for

vaccination but also contributing to the introduction of antibiotics in the 20th century [2, 8].

These advancements have been crucial for controlling the spread of several infectious diseases, including smallpox and polio [1, 8]. They also set the stage for the creation in the late 20th and early 21st centuries of innovative molecular diagnostic technologies, such as polymerase chain reaction (PCR) and next-generation sequencing – markedly increasing the capacity of countries to rapidly identify and genetically characterize pathogens [9-11]. In turn, this has facilitated the formulation of targeted healthcare interventions at an unprecedented pace. An example of this is the sequencing of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) within weeks of identification of the first cases [12, 13], which allowed the development of diagnostics, therapeutics, and vaccines for COVID-19 in the same year of its emergence – significantly enhancing the pandemic response [14, 15].

1.1.3. Spatial analysis to support disease surveillance

In recent years, spatial analysis has come to the fore as a key instrument to explore and understand disease transmission patterns [16, 17]. This type of analysis harnesses geographic information systems (GIS), which are computer information platforms for collecting, managing, and analysing spatial and non-spatial data. Spatial analysis integrates GIS with numerical data, leveraging both statistical and non-statistical techniques, to uncover spatial associations that would otherwise be difficult to notice [16, 17].

Spatial analysis has been used to enhance infectious disease surveillance across various health crises. For instance, it helped public health practitioners visualize the geographic distribution of cases in the 2014–2016 Ebola virus disease (EVD) outbreak in West Africa, enabling the characterization of the role of superspreading and imported cases in sustaining onward transmission of the epidemic, and facilitating the development of targeted interventions [18-21]. Similarly, spatial analysis supported the surveillance of cholera outbreaks in Haiti, allowing the tracking of the spread of *Vibrio cholerae*, the detection of clusters of transmission, and the assessment of the effectiveness of control measures [22-24].

Spatial analysis has also played a vital role in epidemic preparedness and response. During the COVID-19 pandemic, it has been used extensively to study SARS-CoV-2 transmission dynamics, monitor vaccination coverage, and build prediction models [25-30]. These analyses have been crucial for guiding public health strategies and informing policy decisions.

Aside from assisting in outbreak containment, spatial analysis has been used to investigate endemic diseases. For instance, it has been essential for understanding the epidemiology of malaria in sub-Saharan Africa. This includes mapping its prevalence; identifying transmission hotspots (i.e., areas characterized by accelerated and extensive spread of the disease); and understanding the influence of climatic factors on the spread of malaria [31-36]. Evidence from these analyses has guided control efforts, enabling targeted deployments of insecticide-treated nets, vaccines, and other interventions in at-risk areas.

While the studies mentioned above provided critical insights into the spread of diseases, their findings should be interpreted in light of the limitations inherent to spatial analysis. This techniques sometimes relies on oversimplified models or implausible assumptions, which may have limited applicability in the real world [37]. Spatial analysis studies that do not adequately incorporate behavioural, demographic, environmental, and socioeconomic factors shaping disease transmission may also misrepresent the true nature of epidemiological trends, leading to inaccurate findings [38]. Furthermore, analyses based on data aggregated at large administrative units (e.g., nations or regions) may mask local dynamics, which are key for understanding heterogeneities in health outcomes [39]. However, it is important to highlight that extremely complex models may also be impractical for real-time decision-making due to their heavy computational demands alongside the constraints related to the quality and availability of input data. Finally, given that spatial analysis depends on observing patterns, it can uncover associations between variables, but cannot demonstrate causal relationships – restricting the conclusions that can be made from this type of study [37].

Despite these limitations, spatial analysis offers valuable insights for epidemiological investigations by mapping disease patterns, detecting hotspots, and uncovering hidden correlations. When used in conjunction with other methods, it can provide a more holistic understanding of disease dynamics and guide targeted control measures.

1.1.4. Role of social media monitoring in disease surveillance

The advent of social media has paved novel avenues for infectious disease surveillance [40-42]. Platforms such as Facebook and X (formerly Twitter) have emerged as new important sources of data that can help enhance early warning, alert, and response systems. This is mainly due to the near real-time nature of social media data, which allows public health authorities to track initial signs of disease spread at a pace that is beyond the capabilities of traditional surveillance methods [40-42].

By integrating GIS technologies with social media data, researchers can monitor the geographic spread of diseases, enabling the identification of transmission hotspots. For example, Golder *et al*. (2022) [43] analyzed the frequency and content of geotagged X posts from the United Kingdom mentioning terms related to COVID-19, discovering that the volume of posts mirrored disease trends and anticipated case counts up to 14 days in advance of official government reports. Similar findings were observed in Italy, Spain, and the United States [44, 45].

Through the analysis of social media posts, health authorities can also gauge public sentiment towards specific health issues or interventions [46]. In a 2011 study by Salathé and Khandelwal [47], the analysis of X posts relating to the H1N1 influenza vaccine during the second half of 2009 revealed statistically significant correlations between negative sentiments expressed on the platform and lower vaccination rates. In 2022, Ng *et al*. [48] analyzed X posts related to the global outbreak of mpox, discovering three major themes: concerns about personal safety; stigma towards lesbian, gay, bisexual, transgender, and queer communities and racial minorities; and mistrust in the capacity of governmental and public health organizations to control the outbreak.

These types of findings are crucial for identifying people's perceptions, tailoring public health messaging, and counteracting misinformation.

As with spatial analysis, conclusions drawn from social media monitoring must be interpreted taking into account its limitations. Since social media users are generally younger, more educated, and wealthier than other internet users and the wider population [49, 50], it is unlikely that the opinions shared on these platforms are fully representative of real-world trends. This may result in a skewed understanding of public health sentiments and behaviors, leading to public health strategies that fail to meet the needs of overlooked communities. Furthermore, social media accounts may not always use geotags or provide accurate location information in their profiles [51, 52], affecting the representativeness and accuracy of research findings. In particular, low- and middle-income countries (LMICs) are less present on social media partly due to suboptimal internet access [53], while being disproportionately affected by infectious diseases [54]. This makes it even more difficult to correlate online discussions to actual epidemiological changes. Automated sentiment analysis of social media data often fails to capture the subtleties of language, such as sarcasm, irony, and idiomatic expressions, leading to incorrect interpretations of people's opinions and attitudes [55, 56]. Hence, it is essential to incorporate a qualitative component into this type of research in order to obtain deeper insights into the context and nuances behind social media posts, which automated tools may miss, allowing a more accurate interpretation of public attitudes and concerns. Finally, social media monitoring can detect shifts in how frequently a health topic is mentioned. However, it is unable to definitively confirm whether these changes are due to actual infection surges or resulting from increased public interest due to, for

instance, heightened media coverage [57]. Therefore, linking online discussion with epidemiological data from traditional surveillance systems can provide a more comprehensive understanding of how public sentiment translates into health outcomes.

1.1.5. Conclusion

This section has examined the dynamic and ever-advancing field of infectious disease surveillance, highlighting its fundamental role for global health security. The evolution of this field has been characterized by the constant integration of traditional and innovative methods. Emerging technologies like spatial analysis and social media monitoring represent a significant leap forward in countries' capacity to track, understand, and control the spread of diseases. When used in conjunction with other surveillance tools, they can facilitate the detection of disease hotspots, at-risk populations, and emerging public concerns across different areas.

Despite their potential to provide novel public health insights, the use of spatial analysis and social media monitoring for infectious disease surveillance remains limited. This could partly be due to the limitations of these tools and the general lack of familiarity with their use. Addressing this knowledge gap is, therefore, critical for enabling a more systematic application of spatial analysis and social media monitoring in infectious disease surveillance, which could ultimately lead to improvements in the control and prevention of outbreaks. This is especially relevant in the face of relatively understudied diseases like diphtheria, whose global resurgence is causing significant morbidity and mortality [58, 59].

1.2. Diphtheria

1.2.1. Pathogenesis and microbiology

Diphtheria is a highly contagious disease primarily caused by toxigenic strains of *Corynebacterium diphtheriae*, an aerobic gram-positive bacterium characterized by a club-shaped appearance (**Figure 1.1**) [60-62]. *C. diphtheriae* is also known as the Klebs–Löffler bacillus, in honor of the two German bacteriologists, Edwin Klebs (1834–1912) and Friedrich Löffler (1852–1915), who independently discovered the bacterium in the early 1880s [63].

Figure 1.1. Gram staining of *Corynebacterium diphtheriae*

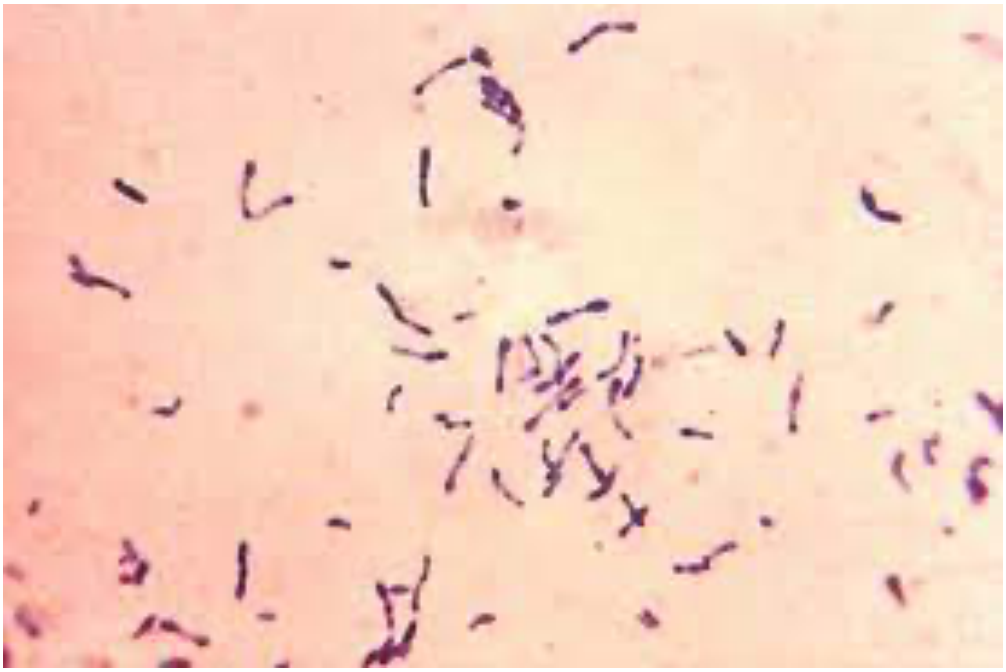


Image courtesy of Dr. T. F. Sellers, Centers for Disease Control and Prevention (CDC) / Public Health Image Library (PHIL), licensed under Creative Commons Attribution 4.0 International (CC BY 4.0) [64]

Four biotypes of *C. diphtheriae* exist: gravis, mitis, belfanti, and intermedius. Although the biotypes have distinctive colonial

morphologies and biochemical parameters, no consistent differences have been observed among them in terms of prevalence or disease severity [65]. Humans are the only reservoir for *C. diphtheriae* [60-62].

C. diphtheriae predominantly colonizes the mucous membranes of the respiratory tract, particularly the pharynx and tonsils [61]. Other common foci of infection include the larynx, nose, and skin. *C. diphtheriae* contains several adhesins, including pili and fimbriae, which mediate adhesion to the host epithelial cells.

Following infection, after an average incubation period of two to five days, *C. diphtheriae* generally secretes the diphtheria toxin – an exotoxin that interferes with the normal functions of human cells by inhibiting intracellular protein synthesis. Ultimately, this process causes cellular damage and local tissue destruction [60, 62, 66]. The diphtheria toxin is responsible for the symptoms and complications associated with diphtheria infection.

The diphtheria toxin can also be produced by two other zoonotic corynebacterial species, *C. ulcerans* and *C. pseudotuberculosis* [67, 68]. Human infection with toxigenic *C. ulcerans* can result in symptoms similar to those caused by *C. diphtheriae*. Infection with *C. pseudotuberculosis* in humans is instead associated with caseous lymphadenitis, a zoonosis of sheep and goats [67].

1.2.2. Clinical characteristics

Transmission of *C. diphtheriae* usually occurs via inhalation or contact with respiratory or skin secretions of an infected person [61, 68]. It is

estimated that an infected individual spreads the disease on average to one to three other people in the initial phase of an outbreak [69]. The average time interval between the onset of symptoms in one diphtheria case and another is around eight days [69]. Both symptomatic and asymptomatic individuals can serve as vectors for the disease. Nevertheless, asymptomatic carriers cause 76% fewer cases than symptomatic individuals [69].

Approximately 31% of those infected remain asymptomatic [69]. Individuals that develop symptoms may present an adherent, thick, greyish-white coating composed of debris, exudate, blood cells, and fibrin in the tonsils and pharynx (**Figure 1.2**) [62, 67]. This "pseudomembrane", as it is often called, is the hallmark feature of respiratory diphtheria. Other symptoms may include malaise, moderate fever, sore throat, and swelling of the neck (**Figure 1.3**) [61, 62].

Figure 1.2. Pseudomembrane covering the tonsils of a diphtheria case

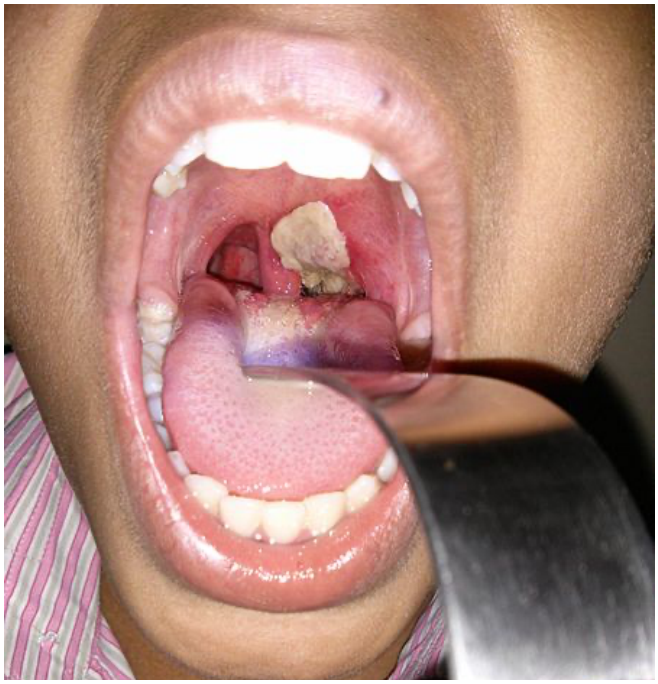


Image courtesy of S. Lal, licensed under CC BY-SA 3.0 [70]

Figure 1.3. Swelling of the neck of a diphtheria case

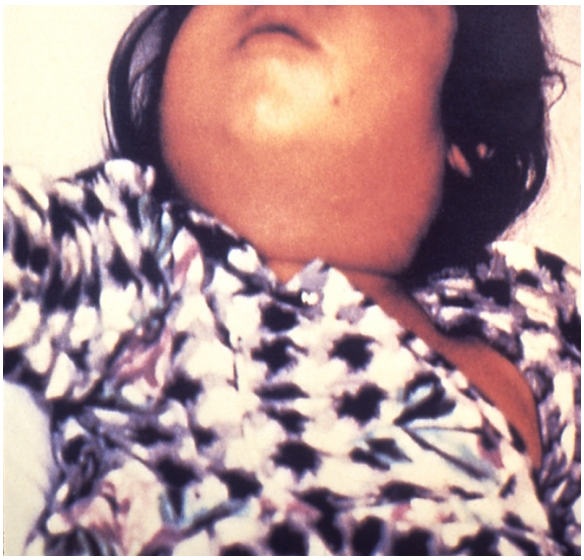


Image courtesy of Dr. H. Smith, CDC / PHIL, licensed under CC BY 4.0 [71]

Diphtheria can also manifest as a skin infection. Cutaneous diphtheria is defined by the development of non-healing ulcers or skin lesions covered by adherent grey membranes (**Figure 1.4**). Although typically milder than respiratory diphtheria, cutaneous diphtheria can still result in severe complications if the infection is not treated or if the bacteria spread to other parts of the body [61].

Figure 1.4. A skin lesion on the leg of a diphtheria case

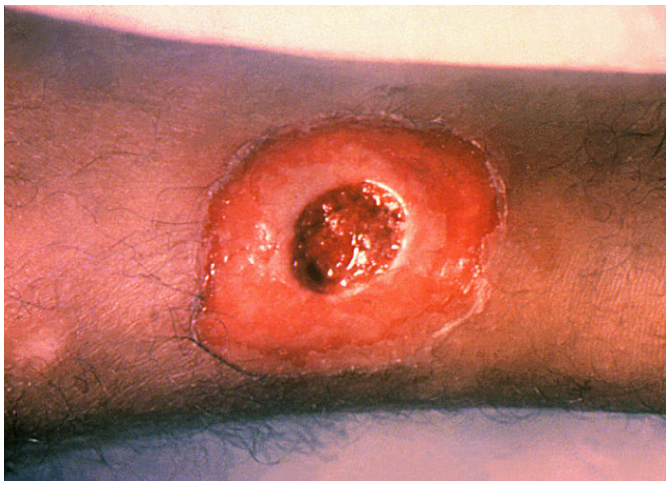


Image courtesy of Dr. W.A. Clark, CDC / PHIL, licensed under CC BY 4.0 [72]

Major complications that can occur following the absorption of the diphtheria toxin into the bloodstream include myocarditis, neuritis, and acute kidney injury [65, 66]. In respiratory diphtheria, the expansion of the pseudomembrane may lead to severe respiratory obstruction. Due to the weakened immune systems, individuals with severe diphtheria may also be more vulnerable to secondary bacterial infections, such as pneumonia [61].

The case fatality rate (CFR) for respiratory diphtheria is generally 5–10% [62]. However, it can reach 29% in individuals who have not received treatment or vaccination [69]. Other factors that can influence the severity and outcomes of the disease include the site of infection and the quantity of toxin produced by *C. diphtheriae* but also pre-existing health conditions, immune status, and age of the host [62]. For instance, children under five years have a 50% higher risk of dying from diphtheria relative to adults aged 20 or older [69]. This may be due to children's immature immune systems and limited prior exposure to *C. diphtheriae*.

1.2.3. Diagnosis

Presumptive diagnosis of diphtheria depends on the observation of clinical signs and symptoms [60, 73]. For patients suspected of respiratory diphtheria, a pharyngeal swab and a nasal swab are obtained [73]. Instead, a swab of the lesion is collected from patients suspected of non-respiratory diphtheria. The specimens are inoculated onto selective growth media, such as tellurite blood agar, for the isolation and cultivation of *Corynebacterium* species [73]. These species exhibit different phenotypes on blood agar plates. For instance, *C. diphtheriae* is

characterized by the presence of grey colonies with a black precipitate (**Figure 1.5**).

Figure 1.5. ***Corynebacterium diphtheriae*** **on tellurite agar**

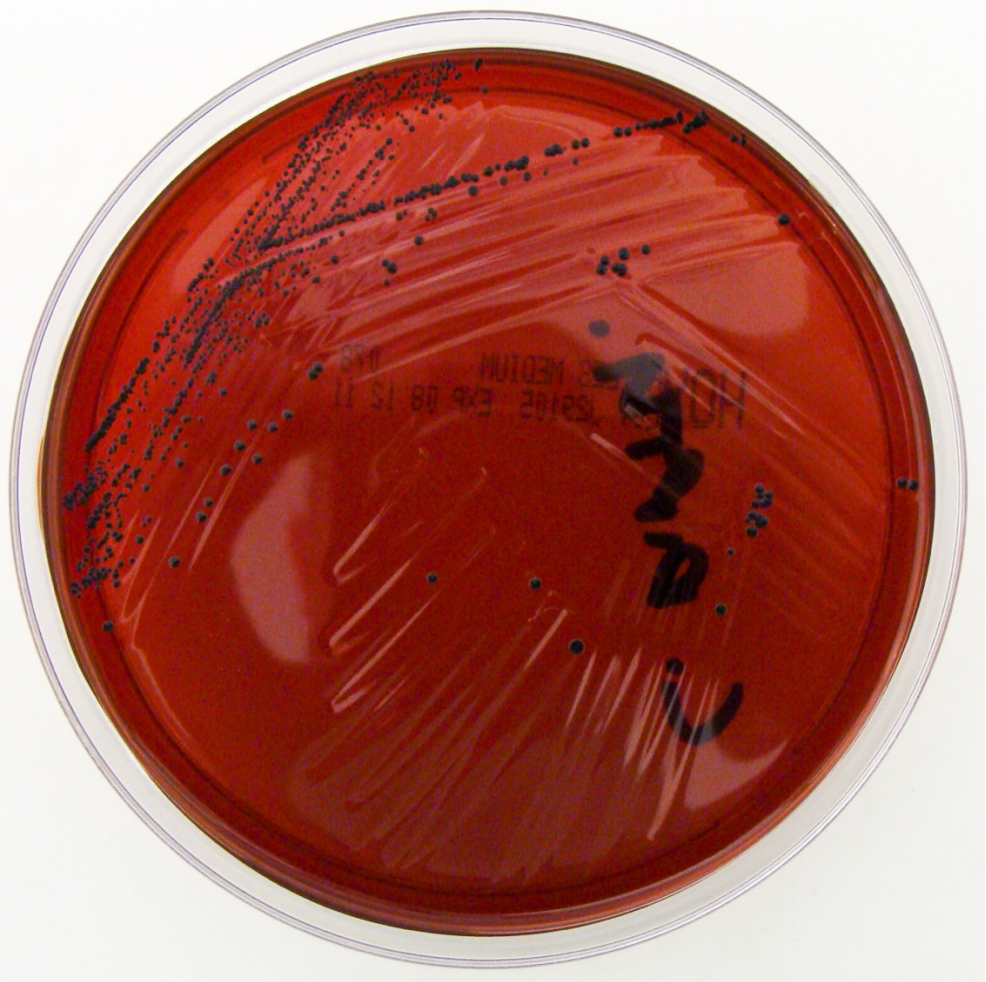


Image courtesy of N. Reading, licensed under CC BY 2.0 [74]

When an organism is identified as a potential *Corynebacterium* species, the isolate is tested to ascertain whether it is toxigenic (i.e., able to produce the diphtheria toxin) using toxin detection assays, such as the Elek test [73]. This in vitro assay measures the presence of diphtheria toxin by evaluating the immunoprecipitation reaction between the toxin and antitoxin antibodies. In the Elek test, toxigenic isolates generally result in visible lines or zones of precipitation, which help to distinguish

them from non-toxigenic isolates. The combination of positive culture and Elek test results confirms the diagnosis [73]. Generally, the two methods require two to five days to produce results [75].

In recent years, molecular assays have become important tools to supplement culture identification and toxigenicity determination. PCR assays target genes or genetic regions specific to *Corynebacterium* species, such as the *tox* gene – which encodes the diphtheria toxin [73, 75]. Amplification of this gene indicates that the strain is toxigenic. Compared to culture, PCR-based methods offer higher sensitivity and specificity as well as faster turnaround times, facilitating early detection of *Corynebacterium* species and rapid initiation of treatment. During large outbreaks, PCR can be used as a standalone diagnostic tool provided that toxigenic diphtheria has been detected by culture and Elek testing in at least five patients [68]. Nevertheless, not all laboratories have the required equipment, reagents, and trained personnel to perform PCR testing, especially in resource-limited settings where diphtheria is endemic.

Surveys conducted in 2017 in the European Union / European Economic Area (EU/EEA) [76] and in 2019 in the Western Pacific Region (WPR) [77] provide important perspectives regarding the global landscape of diphtheria diagnostics. Most countries in both areas have the essential diagnostic capabilities for diphtheria, including culture, biochemical, and toxigenicity testing. However, the EU/EEA has a higher percentage of countries with full reference-level diagnostic capacities, alongside a broader implementation of molecular typing and serological techniques, potentially indicating an overall more advanced laboratory infrastructure compared to the WPR. Both areas present shortcomings

in the training of laboratory personnel; insufficient external quality assurance (EQA) programmes; and difficulties in obtaining the required laboratory supplies, especially for the Elek test and tellurite blood agar. These gaps may have an impact on both the accuracy and timeliness of laboratory diagnosis. It is likely that the challenges observed in the EU/EEA and in the WPR may be more acute in other parts of the world, particularly in LMICs, due to resource limitations, infrastructure inadequacies, and access issues. In turn, this may lead to a greater reliance on clinical diagnosis without laboratory confirmation, potentially resulting in the underdiagnosis or misdiagnosis of diphtheria cases in LMICs.

1.2.4. Treatment

When diphtheria is suspected, treatment should be started immediately, even in the absence of laboratory results, in order to minimize tissue damage, halt disease progression, and decrease the risk of complications [58]. The management of diphtheria necessitates a multi-pronged approach, encompassing supportive care, alongside the administration of the diphtheria antitoxin (DAT) and antibiotics.

Supportive care for diphtheria patients entails a variety of measures meant to reduce symptoms, promote recovery, and prevent complications [62, 78]. These measures may include monitoring vital signs, ensuring adequate hydration and nutrition, administering analgesics, providing respiratory support as required, and ensuring prompt intervention for any potential emerging complication.

DAT, which is made from horse serum, represents the cornerstone of diphtheria treatment [62, 68, 79]. It contains antibodies that specifically target and bind to the diphtheria toxin molecules circulating in the body. This binding helps to prevent the toxin from spreading further and causing additional damage to the patient's tissues and organs [62, 68, 79]. Treatment with DAT leads to a 76% decrease in mortality, with the effectiveness of the medication diminishing the longer its administration is delayed [69]. Given that DAT is an equine product, skin testing must be performed on patients before the injection of DAT [78]. Patients who experience hypersensitivity reactions must be desensitized prior to receiving DAT. Importantly, DAT is in scarce supply worldwide, with regional variations in its availability [76-78], which contributes to inequities in the morbidity and mortality associated with diphtheria.

Finally, therapy with antibiotics like penicillin and erythromycin allows patients to eradicate *C. diphtheriae* from their respiratory tract in about five days, thereby shortening the period of infectiousness by approximately 14 days [69]. This, in turn, reduces the likelihood of further disease transmission [62, 68, 79]. Similarly, macrolides, such as azithromycin and clarithromycin, have demonstrated efficacy in the treatment of diphtheria [80]. Ultimately, the choice of antibiotic is dictated by patient-specific factors, such as allergies and interactions with other medications, as well as local antimicrobial resistance (AMR) patterns.

A 2021 study of 502 *C. diphtheriae* isolates collected from 16 countries across 122 years revealed an increase in the number of AMR genes, especially during the 2010s [81]. Resistance was identified against multiple classes of antibiotics. This trend was not limited to a single

country or region but was observed in both Asia and Europe. This suggests that AMR in *C. diphtheriae* is escalating and geographically widespread. These findings raise concerns about the effectiveness of current diphtheria treatment strategies, highlighting a need to revise existing therapeutic recommendations. Additionally, it is imperative to not only continue but also to actively promote research into new antibiotics and alternative treatment options for diphtheria.

1.2.5. Prevention

The increasing treatment obstacles arising from AMR make diphtheria prevention more critical than ever. The most important intervention for preventing symptomatic disease is vaccination [60, 61, 65]. The first diphtheria vaccine was developed in 1921. Current diphtheria vaccines contain inactivated toxin (toxoid) generally combined with antigens against both tetanus and pertussis (DTP). Other formulations exist, such as DTP in combination with antigens against hepatitis B and *Haemophilus influenzae* type b (the pentavalent vaccine) [58, 65, 66]. For routine immunization, these conjugate vaccines should be given to infants in three doses between six weeks and 18 months of age. Each dose should be separated by at least four weeks [58].

Although completion of the primary vaccination series does not inhibit colonization, it is 87% effective in preventing symptomatic diphtheria [69]. This means that, despite the high vaccine effectiveness, a substantial portion of those vaccinated may still develop symptoms if exposed to *C. diphtheriae*. People who may not acquire full protection include immunocompromised persons and people living with HIV [58].

In the absence of natural boosting, the immunity conferred through primary vaccination gradually wanes, with the share of people with protective antibody levels (≥0.1 IU/mL) decreasing annually by 0.6% following vaccination [69]. Therefore, booster doses are needed to ensure sustained protection, although uncertainty remains regarding the required number and frequency of boosters [58]. Sustaining vaccine coverage of at least 80–85% achieves "herd immunity" – a state in which a sufficiently high portion of a population is immunized against a disease, providing indirect protection to those who are not immune [65].

While vaccines have been highly effective in reducing the burden of many infectious diseases, several factors can hinder or facilitate vaccination. These factors can be broadly categorized into three groups: intent to vaccinate, facility readiness, and community access [82].

Intent to vaccinate refers to the willingness of people to receive vaccination or to have their children vaccinated [82]. Numerous studies have shown that various determinants shape intent to vaccinate, ultimately influencing vaccine uptake. For instance, a systematic review of 64 studies found strong evidence of an association between the uptake of routine vaccines and not considering vaccines as harmful, having a favorable stance toward vaccination, and perceiving few practical obstacles to vaccination [83]. Similarly, a systematic review of 35 publications revealed that perceiving vaccines to be safe was positively associated with the acceptance of the COVID-19 vaccine in sub-Saharan African countries [84]. Additional determinants of intent to vaccinate include knowledge about vaccines, trust in healthcare providers, and cultural norms and community dynamics [82-84]. A major determinant of intent to vaccinate that can lead to vaccine hesitancy is

misinformation. For example, a systematic review of 115 articles found that hesitancy to the measles, mumps, and rubella (MMR) vaccine and other childhood vaccines was concentrated among mothers in the United States with at least college-level education who preferred the internet and social media as sources of health information over healthcare providers [85]. In this review, the prevailing reason for vaccine hesitancy was the fear of autism, which originated from a now widely discredited study published in 1998 that erroneously linked the MMR vaccine and autism following the investigation of 12 children with developmental disorders [86]. However, it is important to interpret findings from the above-mentioned reviews with caution as they emerged from studies that were often based on self-reported data. Despite its usefulness, this type of data may not always perfectly reflect people's actual opinions and behaviours due to social desirability bias, which occurs when survey respondents provide answers that conform to societal norms and expectations [87].

Facility readiness alludes to the capacity of healthcare systems to meet the demand for vaccines [82]. Various studies have investigated the determinants of facility readiness and their influence on vaccination coverage. An analysis of data from the 194 Member States of the World Health Organization (WHO) found that 34% of countries experienced vaccine stockouts at the national level in 2015, often leading to the interruption of vaccination services [88]. This result masks within-country disparities likely complicating the vaccination landscape in these settings. The analysis also revealed that DTP vaccines were the most frequently affected vaccines, accounting for 42% of stockout events. This raises questions on the prioritization accorded to DTP vaccines in national health strategies. The main reasons for the stockouts

were systemic issues, namely government funding postponements (39%), procurement delays (23%), and inadequate forecasting and supply management (18%). Other major vaccination barriers cited in several reviews relate to health workforce gaps, including shortages of healthcare workers (HCWs) alongside inadequacies in their education and training [89-91].

Community access refers to the ability of individuals to utilize and benefit from vaccination services [82]. Multiple systematic reviews have identified geographical distance to healthcare providers as a key barrier to accessing vaccination in both high- and low-income settings [92-94]. Other major deterrents are direct and indirect expenditures related to vaccination [95-97]. Direct expenditures include out-of-pocket payments for vaccination and doctors' consultations, while examples of indirect expenditures are transportation expenses (e.g., fuel, public transport fares) and lost wages (e.g., by taking time off work).

Despite highlighting many of the factors that likely affect vaccination services, the above-mentioned reviews neglected the social determinants of health that can facilitate or hinder vaccine uptake [98]. **Figure 1.6** presents the key determinants of vaccination uptake, including social determinants of health that may influence this vital public health measure. For instance, in areas characterized by political instability, even adequately financed vaccination programmes may fail to reach coverage targets because of healthcare system disruptions and community mistrust in government initiatives. Factors like income inequality, low technological advancements, and inadequate legislation and regulatory frameworks can also significantly impact vaccination uptake. Similarly, people's socioeconomic status, including their

employment situation, housing conditions, and education level, can have an effect on the willingness or ability to be vaccinated. This underscores the importance of cross-sectoral collaborations to improve vaccine coverage levels.

Figure 1.6. Key determinants of vaccination uptake

 **Intent to vaccinate**	• Perception of vaccine safety • Knowledge about vaccines • Trust in healthcare providers • Cultural norms and community dynamics • Misinformation
 **Facility readiness**	• Vaccine availability • Government funding • Procurement timeliness • Supply forecasting and management • Healthcare workforce capacity
 **Community access**	• Geographic distance to healthcare providers • Indirect costs (e.g., travel fees, lost wages) • Direct costs (e.g., out-of-pocket payments)
 **Social determinants of health**	• Political stability • Income inequality • Technological advancements • Legislation and regulations • Socioeconomic status

1.2.6. Outbreak response

The response to a diphtheria outbreak includes multiple key interventions. The early detection and notification of suspected cases is crucial as it enables their prompt isolation and treatment [60, 68, 99]. Mechanisms should be in place for the collection of specimens and laboratory confirmation of diphtheria infection so that individuals who test positive can rapidly receive DAT. Data collected through

epidemiological and laboratory surveillance allows for the detection of changes in disease patterns, identification of transmission hotspots, and evaluation of response interventions [60, 68, 99].

Tracking and tracing close contacts of diphtheria cases is another important intervention as it facilitates the timely detection of additional cases [60, 68, 99]. Close contacts may include family members, fellow students, HCWs, and others who may have been directly exposed to respiratory and skin secretions of a case. To reduce the risk of disease transmission, close contacts may receive prophylactic treatment with antibiotics and booster doses of diphtheria vaccines [60, 68, 99].

Health promotion activities are also critical for raising awareness about diphtheria. Furthermore, during outbreaks, it is essential to strengthen infection prevention and control measures, especially in at-risk settings (e.g., healthcare facilities and schools) [60, 68, 99].

Finally, mass vaccination campaigns targeting affected communities and high-risk groups are vital to increase population immunity and limit further disease spread [60, 68, 99]. It is estimated that vaccination lowers diphtheria transmission by 60%, presumably due to vaccinated individuals experiencing no or milder symptoms that result in less shedding of *C. diphtheriae* [69]. This figure suggests that while vaccination significantly reduces the rate of community spread, a considerable amount of diphtheria transmission can still occur – further underscoring the need to implement the other response interventions mentioned above.

Vaccination campaigns can vary significantly in their design and implementation according to factors such as the epidemiology of the target disease, the demographic characteristics of the intended beneficiaries, and the quality and accessibility of existing healthcare services.

In fixed-site vaccination campaigns, the target individuals are required to bring themselves to existing healthcare facilities, including clinics, medical centers, and pharmacies, to receive vaccines [100, 101]. To be effective, fixed-site campaigns require healthcare services to be relatively accessible. In both LMICs and high-income countries (HICs), this approach has been used for the administration of the seasonal influenza vaccine and more recently for COVID-19 vaccinations [101, 102]. It was also part of the strategy for the delivery of the oral cholera vaccine in Bangladesh and Haiti [103, 104].

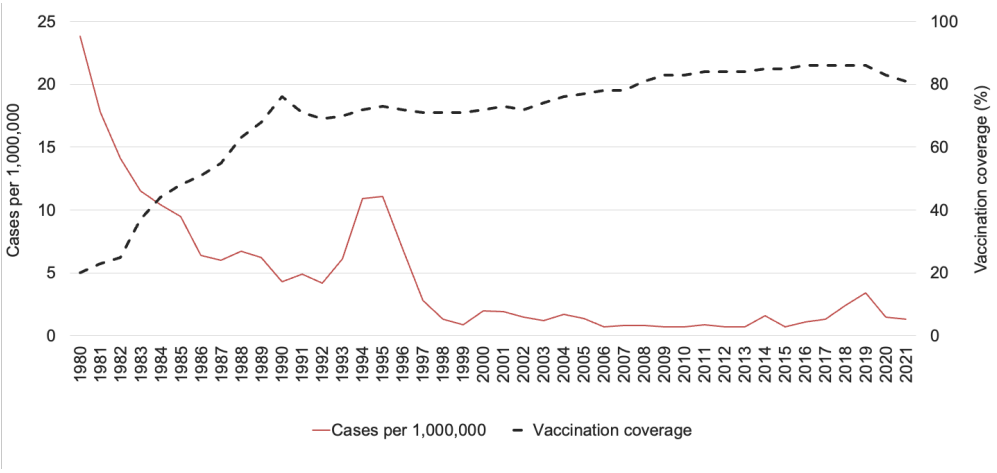
School-based vaccination campaigns leverage academic infrastructures to vaccinate target populations [100, 105]. This approach has allowed large numbers of children and adolescents to be vaccinated against tetanus, diphtheria, and the human papillomavirus (HPV) in countries of all income levels [106-108]. Despite its effectiveness in increasing vaccine coverage among school-age children, a study published in 2020 found that the proportion of WHO Member States implementing school-based vaccination only went up from 58% to 60% between 2008 and 2017 [108]. This could partly be attributed to operational hurdles persisting in areas with low school enrollment or attendance, including some low-income countries (LICs) and rural locations, and in countries with limited financial and human resources for healthcare and education.

Through mobile health units and door-to-door vaccination campaigns, vaccines are delivered near or directly to the target communities, who are often located in remote or hard-to-reach areas where healthcare providers are limited or absent [100, 101]. In low- to high-income countries, mobile units have supported vaccination efforts against several diseases, including COVID-19 [109-111]. Instead, door-to-door campaigns have been crucial in improving the uptake of the polio vaccine in LMICs like Cameroon [112], Ethiopia and India [113].

1.2.7. Global epidemiology

Historically, diphtheria was a major cause of debility, especially among children. In the 1970s, prior to the widespread use of vaccines, about 1 million diphtheria cases occurred each year in LMICs [58]. The reported number of cases decreased dramatically following the launch in 1974 of the Expanded Programme on Immunization (EPI), an initiative led by the World Health Organization (WHO) aimed at achieving universal childhood immunization. Between 1980 and 2000, reported case rates decreased by more than 90% [58, 59] (**Figure 1.7**).

Figure 1.7. Global trends in reported diphtheria case rate (per 1,000,000 population) and DTP3 vaccination coverage, 1980 – 2021



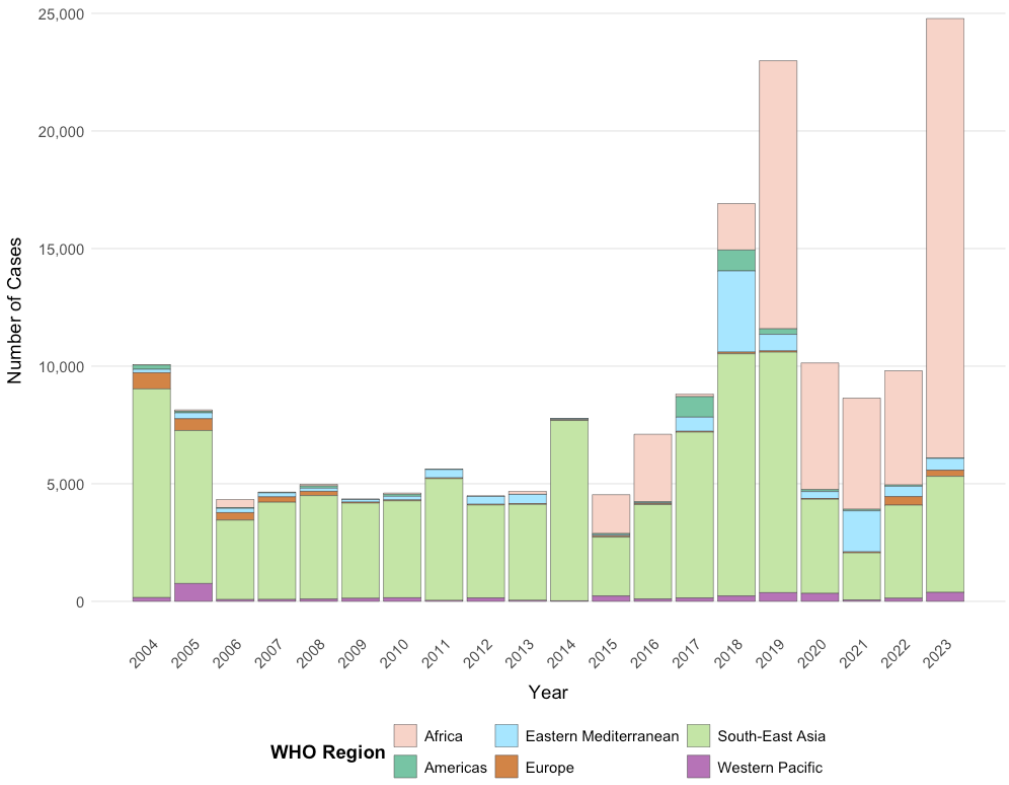
Data source: WHO. Data accessed on 29 February 2024. This figure is based on data obtained from WHO. The interpretation and presentation of the data are the sole responsibility of the author and do not represent the official position of WHO.

The CFR associated with respiratory diphtheria has also reduced over time, going from 52% in the 1880s to 7% in the 1950s. It is likely that this has partly been driven by the development and extensive adoption of medicines and vaccines [69].

Despite the overall reduction in reported cases and mortality, diphtheria remains a major public health issue. Since 2016, there has been a substantial rise in the number of new cases reported worldwide [59] (**Figure 1.8**). Importantly, the lower number of cases reported from 2020 to 2022 should be interpreted with caution as this could partly be due to the restrictions implemented by countries following the emergence of COVID-19. Measures such as lockdowns and social distancing may have limited the spread of infectious diseases like diphtheria by reducing direct interactions between individuals. The same restrictions may have also disrupted public health surveillance, leading to cases of diseases

like diphtheria to be missed. Furthermore, the restrictions may have impacted routine immunization programmes, causing a buildup of susceptible individuals. This, alongside the progressive return to normal social interactions, may have contributed to the significant surge in diphtheria cases observed in 2023, when 24,778 cases were reported – the highest global case count in over 20 years.

Figure 1.8. Number of diphtheria cases by WHO region, 2004 – 2023

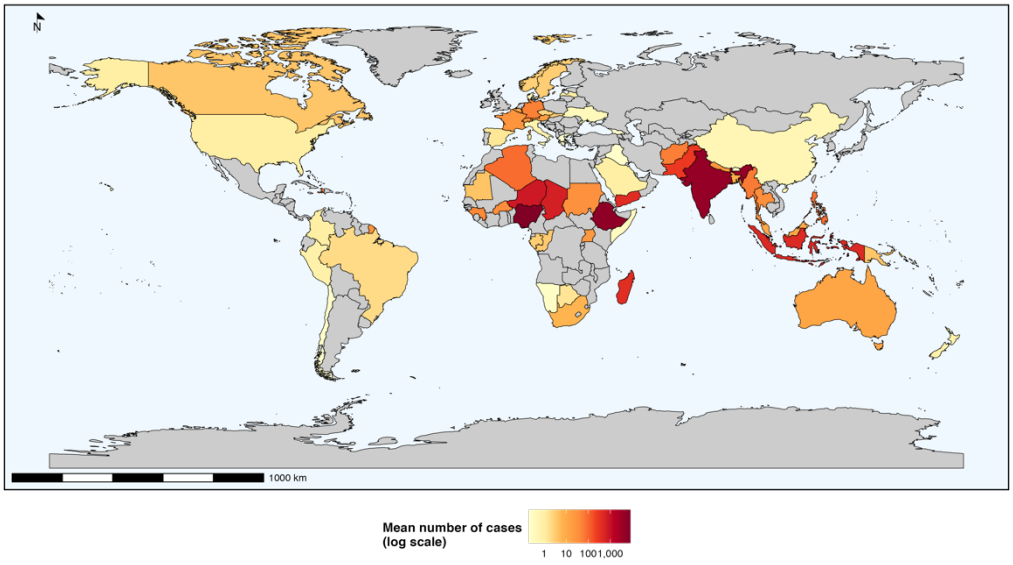


Data source: WHO. Data accessed on 3 August 2024. This table is based on data obtained from WHO. The interpretation and presentation of the data are the sole responsibility of the author and do not represent the official position of WHO.

The bulk of recent outbreaks have occurred in Africa, Eastern Mediterranean, and South-East Asia regions, although diphtheria cases have also been registered in other regions (**Figure 1.9**). Crucially, reported figures are likely an underestimate of the true burden of the

disease as no data were available for several countries. This hypothesis is supported by a 2019 study that reviewed global diphtheria data from 2000 to 2017, finding considerable discrepancies in data reporting and completeness among countries [114]. Moreover, the dearth of diagnostic resources, as noted earlier in this chapter, further exacerbates the underreporting of cases, suggesting that the actual disparity in diphtheria prevalence is probably more pronounced than current data indicate, making it even more difficult to accurately assess the global impact of the disease.

Figure 1.9. Mean number of diphtheria cases by country, 2019 – 2023



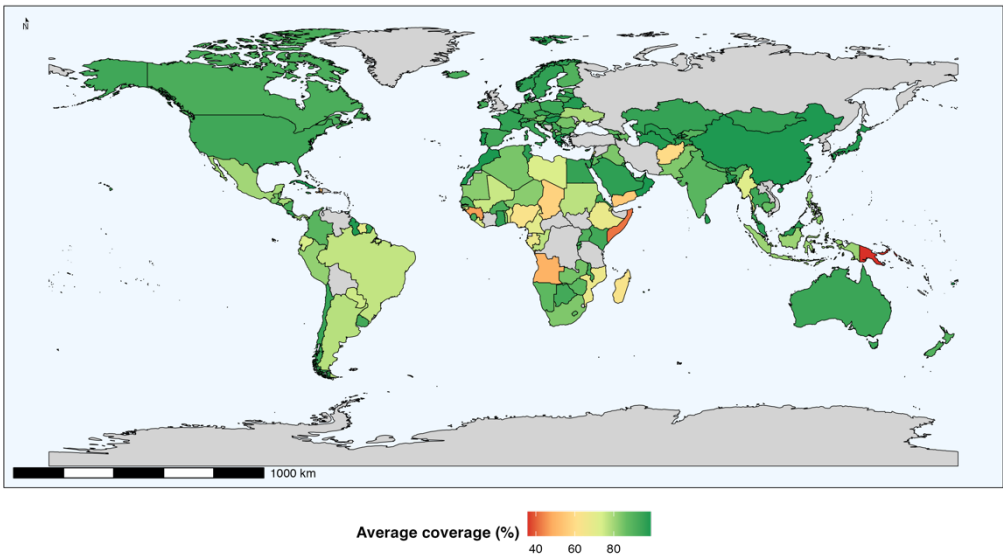
Data source: WHO. Data accessed on 3 August 2024. This figure is based on data obtained from the WHO. The interpretation and presentation of the data are the sole responsibility of the author and do not represent the official position of WHO.

Although medical advancements have successfully contributed to reducing the overall burden and mortality of the disease, some recent diphtheria outbreaks have been characterized by high fatality rates, with the CFR ranging from 3% to 33% in LMICs [69]. This has largely been

attributed to the incorrect diagnosis and inadequate treatment of cases, particularly those detected in the early stages of the outbreaks.

Another factor likely contributing to the global persistence of diphtheria is the suboptimal coverage for the third dose of the DTP vaccine (DTP3) in several areas (**Figure 1.10**). While many countries have achieved near-universal DTP3 coverage, low vaccination rates continue to be reported worldwide, particularly in countries across the African, Eastern Mediterranean, and South-East Asia regions. Differences in healthcare infrastructure and capacities are possibly behind the observed disparities.

Figure 1.10. Global DTP3 coverage among 1-year-olds, 2019 – 2023



Data source: WHO. Data accessed on 3 August 2024. This figure is based on data obtained from WHO. The interpretation and presentation of the data are the sole responsibility of the author and do not represent the official position of WHO.

Analyses of diphtheria data have highlighted demographic differences in the morbidity and mortality associated with the disease. A 2020

systematic review by Truelove *et al.* [69] revealed that the proportion of diphtheria cases aged 20 years and over increased from 17% before 1980 to 36% in the subsequent period. The rise in the percentage of adult cases may be reflective of an increased level of protection in children thanks to improved vaccination coverage, alongside waning protection in adults due to a lack of booster vaccinations.

In terms of gender distribution, the disease occurred equally in males and females in the pre-vaccination era [65]. However, a higher reported case rate was documented in adult women compared to adult men during outbreaks in the 1940s, and a female preponderance was also observed in the Russian Federation alongside other former Soviet Union republics during outbreaks in the 1990s [65]. The vaccination of men during military duty may have contributed to these gender differences.

Socioeconomic factors may also influence the epidemiological trends and patterns of diphtheria. Outbreaks are generally seen in LMICs, particularly among the most economically deprived groups in society who often endure hygienic conditions favorable for the spread of diphtheria alongside other infectious diseases [115, 116]. It is also probable that healthcare factors affect the epidemiology of diphtheria. Areas with weak healthcare systems and infrastructures typically have difficulties in sustaining high vaccination coverage as well as implementing adequate prevention and control measures, as observed during recent large diphtheria outbreaks in Bangladesh, Venezuela, and Yemen [59, 69, 117].

1.2.8. Conclusion

This section has provided an overview on diphtheria, describing core aspects of the disease – from its pathogenesis and microbiology to its outbreak response framework and epidemiology. This comprehensive exploration highlighted how, despite advances in medical science and the existence of an effective vaccine for over a century, diphtheria continues to be a public health threat that causes significant morbidity and mortality worldwide – especially in LMICs.

Despite its resurgence and impact, diphtheria remains relatively understudied. At present, questions persist about the determinants of the observed epidemiological trends and patterns. Previous reviews highlighted some key factors that potentially contribute to the morbidity and mortality associated with the disease, from suboptimal vaccination rates to inadequate healthcare infrastructure and surveillance capacities. Nevertheless, these analyses seem to have overlooked other elements that may increase the likelihood of disease, including behavioural, environmental, health-related, and socioeconomic determinants. Furthermore, none of these reviews provided measures of association between the potential risk factors and diphtheria.

Identifying the risk factors for diphtheria and estimating the measures of association with the disease are fundamental steps to better understand the roots behind the transmission and persistence of diphtheria. This information is essential for developing effective public health strategies and policies against the disease. Therefore, research in this area is needed to bridge the existing knowledge gap and support the implementation of more effective interventions.

What is also missing from the literature are studies that analyze people's knowledge and views regarding diphtheria, particularly in countries with active outbreaks or where the disease remains a persistent threat. Investigating these topics is crucial given that public perception and understanding of a specific health issue inevitably affects the uptake of preventive measures like vaccination. Ultimately, such research can help identify existing communication gaps and design more effective public health strategies.

1.3. The case of Haiti

1.3.1. Rationale for researching diphtheria in Haiti

Among the countries grappling with diphtheria, Haiti stands out as one of the most severely affected by the disease. The prevention and control of the outbreak in the country is complicated by distinctive circumstances and compounding challenges, including vulnerability to natural disasters, concurrent health crises, and civil unrest – all of which considerably strain an already under-resourced and overburdened national healthcare system [118]. In particular, Haiti's sociopolitical and economic situation has experienced a marked deterioration in the past few years, following the assassination of President Jovenel Moïse by unknown assailants in July 2021 [118]. Thus, the country represents a compelling case study for investigating how various factors converge to shape the transmission and persistence of diphtheria.

Information on the epidemiology of diphtheria in Haiti is limited. Data from the Pan American Health Organization / World Health Organization (PAHO / WHO) pointed to a resurgence of diphtheria transmission in the country starting in 2014, with 1,750 suspected cases

reported as of July 2024 [119]. Among the 470 confirmed cases, there were 96 deaths, resulting in an overall CFR of 20.4%. A report from PAHO / WHO mentioned insufficient levels of vaccination as potential contributors to the observed trends, with the national coverage for the fourth dose of the DTP vaccine (DTP4) being less than 50% – far below the 95% coverage recommended by PAHO / WHO's Regional Immunization Action Plan [120]. Furthermore, according to the report, the country did not have a national vaccination policy for healthcare workers, while suspected cases were not systematically vaccinated [120].

The potential causative link between suboptimal vaccination and diphtheria in Haiti was also highlighted in two research studies conducted during the most recent outbreak [121, 122], which found that most of the fatal cases had not been vaccinated or had an unknown vaccination history (69–95%). Additionally, a high CFR (31–50%) was found among confirmed cases in both studies. The elevated rate of mortality was attributed to the late identification of cases, delays in the administration of medications, and inadequate healthcare [122].

Collectively, these papers offer valuable evidence on the clinical characteristics of diphtheria cases in Haiti, laying a crucial foundation for the development and implementation of targeted interventions. Nevertheless, they also pointed towards a need for a more in-depth description of the epidemiological patterns of the disease in the country. Furthermore, the interplay between diphtheria and the factors contributing to its spread remains largely underexplored. Additionally, the perceptions, attitudes, and opinions of the Haitian population towards the outbreak, vaccination, and other implemented public health measures are unknown.

This knowledge is vital for acquiring a deeper understanding of diphtheria dynamics in Haiti alongside designing innovative interventions that address the root causes of its resurgence and transmission. Several domains of public health could potentially benefit from such research, including infectious disease surveillance, vaccination, health promotion, and outbreak response. The generated insights and lessons will likely transcend geographical boundaries, informing response strategies to other diseases that disproportionately affect the world's most vulnerable communities.

1.3.2. Initial research plan

My research project initially aimed to disentangle the complexities surrounding diphtheria in Haiti. The motivation for this project came from seeing firsthand the impact of the disease in the country, where I worked as an epidemiologist for PAHO / WHO in 2018. My time in Haiti was a pivotal moment for me as it offered a unique perspective into the challenges of responding to outbreaks in areas with fragile healthcare systems and profound socioeconomic vulnerabilities. This experience also sparked in me the desire to delve deeper into the specific dynamics of diphtheria transmission in Haiti and identify potential solutions for its prevention and control.

As originally envisioned, the project consisted of three interconnected studies. The first study was a systematic review and meta-analysis of risk factors for diphtheria, which had been prompted by the absence of a comprehensive analysis on this subject. The intention was to produce holistic insights into the factors contributing to diphtheria infection and

generate summary estimates of the association between investigated exposures and the disease.

The second study was a spatial analysis of the epidemiology of diphtheria in Haiti. The aim was to leverage GIS technologies to analyze the geographic and temporal distribution of cases in the country, detect hotspots of transmission, and identify potential factors associated with the reported case rate of the disease. The time spent in the country allowed me to become familiar with the epidemiological data collected by Haiti's Ministry of Public Health and Population (Ministère de la santé publique et de la population; MSPP) and facilitate the process of acquiring access from national authorities to a comprehensive diphtheria dataset for the period of December 2014 to June 2021.

The third study was originally conceived as a qualitative investigation based on interviews with various stakeholders, including senior officials from the MSPP, healthcare professionals working in diphtheria hotspots in Haiti, and staff from health organizations present in the country. Specifically, the study set out to explore the experiences and views of these stakeholders regarding the barriers and facilitators of diphtheria prevention and control in Haiti. Furthermore, the study sought to investigate the perceived impact of recent events, including the emergence of COVID-19 and the country's sociopolitical crisis, on diphtheria response efforts.

It was anticipated that this mixed-methods approach would have provided a detailed picture of the epidemiology of diphtheria in Haiti and the determinants of its persistence. As such, the research project sought to offer robust evidence that could inform the development of

targeted public health strategies for diphtheria and other vaccine-preventable diseases, while uncovering current knowledge gaps and guiding future research.

1.3.3. Changes to the third study

The first two studies of the research project were completed during the first two years of my PhD programme. However, subsequently, considerable hurdles arose that prevented the realization of the third study as initially planned, despite completing the research protocol and obtaining clearance from the University of Nottingham's School of Medicine Research Ethics Committee (**Appendix 1**).

As mentioned earlier, as part of the study, interviews were going to be conducted with public health stakeholders in Haiti. These were going to be chosen based on their knowledge of and involvement with the diphtheria response in the country. Participants were going to be selected purposively using my professional network of contacts in Haiti and through a search of the published and grey literature. To facilitate discussions with the participants, an interview guide with open-ended questions was going to be used.

Originally, the interviews were meant to be carried out in person. However, due to Haiti's fragile security situation, it was collectively decided that all interviews would be conducted remotely either on the phone or online using Microsoft Teams on dates and hours most suitable to the participants. All interviews were expected to last up to 60 minutes. The language used was going to be either English or French, according to the participants' preferences. After receiving permission from the

participants, the interviews were going to be audio recorded. Phone interviews were going to be recorded using an audio recorder and saved in MP3 format. Online interviews were going to be recorded using Microsoft Teams' recording feature and saved in MP4 format. All recordings were going to be stored on the University of Nottingham's online network, which was going to be accessible only to the research team.

Data analysis was going to start once the first interview had been completed and would have continued until a theory of diphtheria persistence in Haiti had been developed. The analysis was going be conducted using Grounded Theory, an inductive research method in which theories are built from the ground up through the continuous collection and analysis of data, rather than the testing of preconceived theories [123, 124]. At the end of the analysis, a tentative model of the perceived barriers and facilitators of diphtheria prevention and control in Haiti alongside the perceived impact of recent events on the diphtheria response would have been elaborated. The model would have been presented in a diagram illustrating the main emerging themes, and the links between them. The model would have been accompanied by a narrative synthesis, with relevant supporting quotes, to enhance understanding and demonstrate the grounding of the findings in the data.

Participants were going to be recruited until theoretical saturation, which refers to the point when collecting new data yields no further insights about the emergent theory [123, 125]. Accordingly, the number of participants was not predetermined, although it was anticipated that fewer than 30 participants were likely going be sufficient to reach

saturation based on past studies that used similar research methods [126, 127].

Over time, however, it became clear that conducting such a study would have been extremely challenging. Although the study was going to be carried out remotely, it was highly unlikely that potential participants would have been able (or willing) to participate in phone interviews, given the competing priorities caused by Haiti's continuous sociopolitical unrest and violence.

This realization required a reassessment of the research approach and the exploration of alternative methods. After discussions among members of the research team, it was decided that the third study would involve analyzing posts relating to diphtheria in Haiti shared on the social media platform X between December 2014 (when the latest outbreak began) and December 2022. It was postulated that examining these tweets might have helped understand how the public perceived the outbreak and the response activities implemented by the MSPP.

After receiving ethical clearance (**Appendix 2**), X posts started being gathered as part of the study. However, during the data collection process, it emerged that X had three levels of access to their data: Basic; Elevated; and Academic, with Academic level providing access to the largest amount of data. Unfortunately, it was only possible to obtain Elevated access, which only allowed the retrieval of posts from the preceding 7–10 days. Furthermore, none of these posts had geographical coordinates, making it challenging to map and contextualize the information reported by X users.

Through the personal connections of one of the members of the supervisory team (DB), contacts were established with researchers at Boston Children's Hospital (BCH), who are affiliated with Harvard University and who have access to all historical posts through a prior arrangement with X. These researchers agreed to set up a data sharing agreement with the University of Nottingham, enabling the collection of historical tweets about diphtheria. This agreement helped solve the issue of accessing large volumes of tweets, while complying with X's data privacy policies and regulations.

Given that only a few posts from Haiti could be retrieved, the focus of the analysis was expanded from a national level to a global scale. This was in line with the scope of the systematic review, which also adopted a worldwide perspective to examine risk factors for diphtheria. The study period too was extended, encompassing over ten years, from 1 January 2012 to 31 December 2022. It was hypothesized that the broader scope and longer timeline would enable a more comprehensive and nuanced understanding of the public discourse surrounding diphtheria, with insights relevant for a range of settings and contexts.

1.3.4. Conclusion

This section has explored the rationale behind researching diphtheria in Haiti. The original research plan had three components: a systematic review and meta-analysis to identify the risk factors for diphtheria, a spatial analysis to investigate the epidemiology of the disease in Haiti, and a qualitative study consisting of in interviews with key informants to understand the barriers and facilitators of diphtheria prevention and control in the country. This mixed-methods approach was intended to

provide a holistic view of the epidemiological landscape and the determinants of diphtheria persistence in Haiti.

The third study was significantly adapted in response to the deterioration of the sociopolitical situation in Haiti. The research methods shifted from in-person to online interviews to an analysis of X posts about diphtheria. Due to unforeseen circumstances, the focus of this analysis was also broadened from concentrating solely on Haiti to including worldwide posts. Although these changes were unplanned, they allowed to capture a wide spectrum of public perceptions relating to the disease, generating global insights that could still reflect and inform the local understanding of diphtheria in Haiti and other countries experiencing similar outbreaks.

The evolution of the methodological approach alongside the various modifications required throughout this project highlighted numerous lessons, including the need for adaptability in research – particularly when conducting studies related to volatile settings. This and other valuable insights gained during the last six years will be thoroughly explored in the discussion chapter of this thesis.

1.4. Thesis objectives and structure

1.4.1. Thesis objectives

This research project sought to fill the aforementioned knowledge gaps by exploring the drivers of global diphtheria persistence. As such, it aimed to address a critical question: what are the underlying factors contributing to diphtheria persistence?

To answer this question, three research objectives were set:

1. To identify the risk factors for diphtheria and, when possible, estimate the strength of association between these factors and the disease.
2. To characterize the spatial epidemiology of diphtheria in Haiti from December 2014 to June 2021.
3. To investigate the global discourse surrounding diphtheria on X between January 2012 and December 2022.

1.4.2. Research approach

To achieve the objectives of this research project, multiple techniques were employed, namely systematic review and meta-analysis (objective 1), GIS and spatial analysis (objective 2), and descriptive and statistical methods alongside Grounded Theory (objective 3). This decision was based on the realization that a complex phenomenon like the persistence of diphtheria could not have been investigated using a single methodology.

The quantitative aspects of this research adhered to the positivist paradigm, which traces its origin to the early 19th century [128, 129]. Positivism is a philosophical approach that assumes the existence of a single, objective reality governed by fixed nature laws that can be understood only through empirical observation and measurement. The employed quantitative techniques allowed the analysis of large-scale data, identifying patterns and trends, as well as measuring associations between diphtheria and potential risk factors.

Conversely, the qualitative aspects of this research were informed by the interpretivist paradigm, which has its roots in the late 19th century [130, 131]. Interpretivism is a philosophical approach that presupposes the coexistence of multiple social realities constructed by people's personal experiences, cultural backgrounds, and surrounding conditions. According to this philosophy, the understanding and interpretation of the same phenomenon can vary from one individual to another. In line with these principles, the qualitative component of the research provided insights into public perspectives, opinions, and concerns regarding diphtheria and governments' response efforts.

The combination of positivism and interpretivism, often referred to as pragmatism, is particularly well-suited for this research [132, 133]. The pragmatist paradigm, which originated in the late 19th century, sees reality as dynamic and continuously evolving based on human experience and practical consequences of ideas. This view of reality directly influences pragmatists' research focus, which is centered on solving real-world problems. Rather than adhering to abstract principles or rigid theories, pragmatists are open to adopt multiple methods and different perspectives to produce knowledge that can improve practice. In this research, the adoption of pragmatism and the use of a mixed-methods approach allowed to pinpoint some of the factors contributing to diphtheria persistence and identify potential interventions that could be implemented to improve the control and prevention of the disease.

1.4.3. Thesis structure

This thesis consists of five chapters, three of which describe research papers that are either published or submitted for publication in peer-reviewed journals.

Chapter 1 has provided an introduction on the area of research, describing the current knowledge gaps, alongside presenting the objectives and approach of this research.

Chapter 2 describes the systematic review and meta-analysis of risk factors for diphtheria. The chapter outlines the search strategy and the methods used to assess the quality of the studies and the strength of the evidence. It then reports on the identified risk factors.

Chapter 3 gives an overview of GIS and spatial analysis and describes the specific techniques used for the study. The chapter then provides contextual details regarding Haiti. Finally, it presents the subnational distribution of confirmed cases in the country, the hotspots of disease transmission, and the identified factors potentially associated with the reported diphtheria case rate.

Chapter 4 offers background information on the use of social media in public health research. Additionally, it illustrates the methods used for the analysis based on X data and reports study results, focusing on the volume and characteristics of posts related to diphtheria, and presenting the underlying themes and subthemes in the data.

Chapter 5 summarizes the results of the three studies, explaining how these are interrelated and fit into a cohesive theoretical model. The chapter also provides recommendations for future public health practice and research. Finally, it discusses the limitations and strengths of the thesis.

Appendices include additional materials that support the thesis, including ethical approval documents, tables, and figures that supplement the results presented in the main chapters.

Chapter 2. Risk factors for diphtheria: A systematic review and meta-analysis

2.1. Introduction

Diphtheria remains a formidable public health threat, especially in LMICs. As discussed in Chapter 1, past literature reviews [58, 69, 114] suggested that the persistence of the disease is likely due to a multitude of factors. Prominent among these factors is inadequate vaccination coverage. Areas with suboptimal coverage tend to have large portions of the population vulnerable to diphtheria, which puts these areas at risk of experiencing outbreaks and sustained disease transmission [69, 114, 115]. As evidence of the importance of vaccination, an analysis of surveillance data reported annually to WHO and UNICEF found that 78% of diphtheria cases detected globally between 2000 and 2017 were either unvaccinated (65%) or partially vaccinated (13%) [114].

While highlighting some key potential risk factors for diphtheria, past literature reviews overlooked several elements that may increase the likelihood of disease, including various behavioural, environmental, health-related, and socioeconomic determinants. Furthermore, none of these reviews provided measures of association between the identified risk factors and diphtheria.

This chapter presents a comprehensive systematic review of diphtheria risk factors and meta-analysis of the strength of association between

these factors and the disease. In this study, a "risk factor" was defined as an element or a variable that can change the probability of diphtheria occurrence. To address the limitations of previous reviews, this study aimed to identify all risk factors for diphtheria, assessing the degree of association with the disease – whenever possible. Therefore, the present research was guided by two questions: what are the risk factors for diphtheria? What is the strength of association between these factors and the disease?

Subsequent sections of this chapter illustrate the methods used for the systematic review and meta-analysis. The main research findings are then presented and discussed within the wider diphtheria literature. Finally, policy and practice recommendations are provided based on evidence from this study, alongside suggestions for future research on diphtheria risk factors.

2.2. Methods

This systematic review and meta-analysis was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement [134] (**Appendix 3**). Before starting the study, the protocol was registered with PROSPERO (CRD42019128741) [135]. The authors received no specific funding for this study.

2.2.1. Search strategy

A literature search was conducted using EMBASE [136], MEDLINE [137], PubMed [138], and Web of Science [139]. These databases were chosen for their robust search capabilities and extensive range of biomedical and health-related literature, including peer-reviewed

articles, preprints, and book chapters. The Gavi publications library [140], OpenGrey [141], and WHO Library Information System (WHOLIS) [142] were also consulted as they collectively house a broad spectrum of grey literature, from technical and research reports to doctoral dissertations and conference papers. By incorporating both published and grey literature, this approach ensured a thorough and unbiased review, thereby enhancing the overall reliability and validity of the findings.

All databases were searched from inception until January 2023. No restrictions were applied. The search strategy combined Medical Subject Headings (MeSH), free-text terms, and keywords related to diphtheria and risk factors. The strategy was developed for MEDLINE and adapted for EMBASE, PubMed, and Web of Science (**Appendix 4**). Only the term "diphtheria" was used for the other databases. Native speakers translated studies in non-English languages.

2.2.2. Eligibility criteria

To be included, studies had to meet two criteria: report on diphtheria cases, as described by the authors – including individuals with a laboratory or clinical diagnosis of diphtheria [68]; and, present estimates of association for at least one potential risk factor or sufficient data to calculate these. Eligible measures of association were odds ratios (ORs), risk ratios (RRs), alongside regression and correlation coefficients.

Studies that focused on non-modifiable risk factors for the disease (e.g., age, sex, and ethnicity) or on risk factors for diphtheria severity or mortality were deemed to be outside the scope of the review. Single case

reports, qualitative studies, commentaries, editorials, economic analyses, letters, literature reviews, and animal studies were also excluded.

2.2.3. Study selection

A single reviewer exported all studies to EndNote X9 (Clarivate Analytics, Philadelphia, USA). The study selection then followed three stages. Firstly, one reviewer screened the titles of all papers to eliminate those that were clearly irrelevant to the review. Secondly, two reviewers independently screened the abstracts of retained papers, excluding those that did not meet the eligibility criteria. Thirdly, both reviewers independently examined the full text of each study deemed potentially relevant based on the abstracts. References of included studies were screened to identify other relevant publications. Disagreements were resolved by discussion between the two reviewers and, when needed, referred to a third person.

2.2.4. Data extraction

Two reviewers independently extracted data from included studies using a standardized Excel spreadsheet (Microsoft, Redmond, United States). For each study, the following variables were recorded: last name of the first author, year of publication, country where the study was conducted, study design, number of cases, number of controls, definition of cases, definition of controls, age range of the sample, diagnostic method of diphtheria infection, considered confounders, data collection technique, and effect estimates. If studies did not report effect estimates, available data in those papers were used to calculate crude ORs from two-by-two tables. Where multiple estimates were provided,

those adjusted for the most confounders were selected. The fully adjusted estimates were the most homogenous in terms of included confounders. This choice was made to reduce the impact of confounding as a source of heterogeneity.

A sensitivity analysis was conducted to evaluate the impact of adjustment for confounding on the identified risk factor with the most evidence (i.e., incomplete vaccination). A sensitivity analysis was also performed to assess the impact of excluding one preprint article on the pooled estimate for another risk factor (i.e., contact with a diphtheria case).

2.2.5. Quality assessment

Two reviewers independently evaluated the methodological quality of case-control, cohort, and cross-sectional studies using the Newcastle-Ottawa Scale (NOS) – a widely adopted tool for assessing the quality of observational studies during systematic reviews [143, 144]. Alternative tools were considered, such as the Critical Appraisal Skills Programme (CASP) Checklists [145], the Downs and Black Checklist [146], the Joanna Briggs Institute (JBI) Critical Appraisal Tools [147], and the Risk of Bias in Non-randomized Studies - of Interventions (ROBINS-I) [148]. However, the NOS was ultimately deemed to be the most suitable instrument for this systematic review because of its comprehensive evaluation criteria, ease of application, and established validity and reliability [143, 144].

The evaluation was based on three main domains: selection of study groups, comparability between cases and controls, and ascertainment of

the outcome of interest. For case-control and cohort studies, eight items were assigned one or two points, for a maximum score of nine. For cross-sectional studies, seven items received one or two points, for a maximum score of 10. Each score denoted a different level of quality: low (0–3), moderate (4–7), and high (≥8).

The quality of evidence for each risk factor across the studies was assessed using the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) criteria – a popular tool for appraising the robustness of evidence and the strength of healthcare recommendations [149, 150]. Other tools were explored, including the strength of evidence approach developed by the Evidence-based Practice Center (EPC) program of the Agency for Healthcare Research and Quality (AHRQ) [151] and the Oxford Centre for Evidence-Based Medicine (OCEBM) Levels of Evidence [152]. Nevertheless, the GRADE criteria were preferred due to their methodical and thorough evaluation framework, applicability across study designs, and endorsement by leading institutions like Cochrane [153] and WHO [154].

In line with the GRADE approach, the quality was rated very low, low, moderate, or high [149, 150]. Evidence from observational studies was initially rated by default as low quality due to residual confounding, which occurs when unaccounted or inadequately measured variables continue to distort study results despite attempts to control for them [153]. Evidence could then be downgraded based on the consideration of five domains: risk of bias (as indicated by the NOS score), imprecision (if the confidence intervals were wide or the total population size was less than 400 – a threshold "rule-of-thumb" value), inconsistency (if there was variability in the direction or magnitude of effect across

studies), indirectness (if there were factors relating to the study population that could limit generalizability), and publication bias (if there was a high probability of unreported studies). Criteria for upgrading evidence included a dose–response relationship or large effect size (OR≥2 or OR≤0.5).

While the exact number of disagreements regarding the NOS and GRADE assessments was not formally recorded, differences in the interpretation of study quality and evidence were infrequent and generally resolved through collaborative discussion. When consensus could not be reached, a third reviewer provided the final decision.

2.2.6. Data analysis

The included studies were synthesized qualitatively to present an in-depth narrative description of the findings and to highlight recurring patterns. As diphtheria is a relatively rare event, ORs and RRs were considered equivalent. This equivalence is based on the statistical principle that when an outcome is rare (typically below 10%), the difference between ORs and RRs is negligible, making them nearly interchangeable [155, 156]. This approach facilitates the integration and comparison of results from various studies, ensuring the evidence is presented coherently and consistently.

A meta-analysis was conducted if two or more studies captured the same risk factor but from different samples. The random effects model was chosen for the analysis due to the expected variations in study methods and populations [157]. Original data from the studies were combined to calculate pooled odds ratios (PORs) with 95% confidence

intervals (CIs) using the Mantel-Haenszel method, which stratifies data, calculates and weights odds ratios for each stratum, and controls for confounding variables [158]. If studies only presented adjusted data, the generic inverse weighted method was used, which assigns weights to each study's effect estimate based on the inverse of its variance – thereby prioritizing studies with greater precision [159]. The weighted estimates are combined to generate a pooled effect estimate, using its variance to calculate the 95% CI.

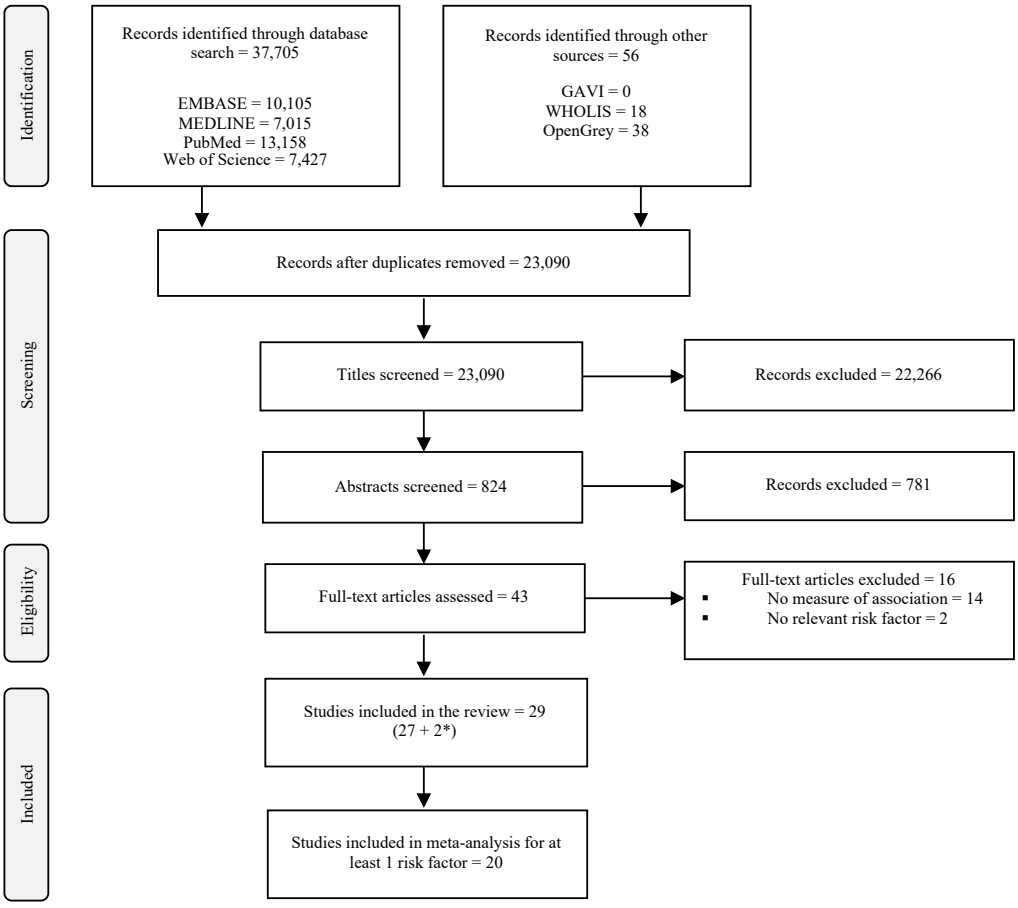
Other methods that could have been used for the meta-analysis include the Bayesian hierarchical model and meta-regression; while the former is complex and computationally demanding, the latter necessitates large datasets and is not primarily designed to generate an overall effect estimate [160, 161].

To facilitate the comparison of the study results and their contributions to the overall estimates, forest plots were produced. These are charts that display the effect estimates, confidence intervals, and weights of individual studies [162]. The effect estimate is depicted as a square or dot, confidence intervals are represented by horizontal lines, while the pooled estimate is displayed as a diamond. Heterogeneity among studies was assessed using the I^2 statistic, a descriptive measure that quantifies how much of the variation in study results is due to actual differences rather than chance [163]. Values of 25%, 50%, and 75% revealed low, medium, and high heterogeneity, respectively. All analyses were performed using Review Manager 5.3 (Cochrane Collaboration, Copenhagen, Denmark).

2.3. Results

After the initial search, 43 publications were selected for full-text review (**Figure 2.1**). Fourteen were excluded as they did not provide measures of association and two because the investigated exposures were irrelevant. In addition to the remaining 27 publications, two other references were identified by hand searching. Twenty-nine papers were included in the review and 20 in the meta-analysis. Nine papers were excluded from the meta-analysis as they reported on risk factors that had not been investigated in other studies. Therefore, their results could not be pooled with data from other publications.

Figure 2.1. Search and selection for the review and meta-analysis



*Additional papers included following hand searching of reference lists.

2.3.1. Study characteristics

Table 2.1 describes the characteristics of the 29 included studies. Overall, there were 14 case-control studies, eight cross-sectional studies, six ecological studies, and one cohort study. The studies covered a period from 1907 to 2021, with most of the studies (55%) focusing on outbreaks that occurred prior to the year 2000. Only one study reported on an outbreak with cases identified both before and after 2000. The studies were conducted in five WHO regions: Europe (nine), Americas (eight), South-East Asia (six), Eastern Mediterranean (three), and Western Pacific (three). No study was done in Africa. Notably, the only study from the Americas addressing diphtheria in Haiti was the spatial analysis performed as part of this research project (Chapter 3).

Table 2.1. Characteristics of studies included in the systematic review

Reference	WHO region	Country	Study period	Cases (n)	Controls (n)	Age (years)	Method of diagnosis	Statistical analysis	Considered confounders	NOS score
Case-control studies										
Allam 2016 [164]	SEAR	India	2013	63	63	10-45	Clinical	Multivariate logistic regression	Age, day of healthcare visit, location	7
Bisgard 2000 **[165]**	EUR	Russia	1991-1992	217	2,168	0-14	Clinical or laboratory	2x2 cross-tabulation	Age, location	8
Bitragunta 2008 **[166]**	SEAR	India	2003-2006	123	123	0-10	Laboratory	Conditional logistic regression	Age, location	7
Brennan 2000 [167]	EUR	Russia	1995-1996	39	117	40-49	Laboratory	Conditional logistic regression	Age, location	9
Chen 2018 **[168]**	EUR	Ukraine	1992	262	517	0-14	N/A	2x2 cross-tabulation	Age, location	7
Faria 1971 **[169]**	AMR	Brazil	1969	27	24	1-12	N/A	2x2 cross-tabulation	Age, location, sex	4
Husada 2018 **[170]**	SEAR	Indonesia	2011-2015	27	108	0-18	Laboratory	Multivariate logistic regression	Age, location	6
Jones 1985 **[171]**	EMR	Yemen	1981-1982	47	94	N/A	Clinical or laboratory	Multivariate logistic regression	Age, location, sex	6
Murakami 2010 **[172]**	WPR	Vietnam	2005-2006	88	352	1-32	Clinical	Multivariate logistic regression	Age, location, sex	7

Nassar 2021 **[173]**	EMR	Yemen	2019	76	152	N/A	Clinical or laboratory	Multivariate logistic regression	Age, location, sex	8
Quick 2000 **[174]**	EUR	Georgia	1995-1996	218	408	0-75	Clinical	Multivariate logistic regression	Age, location, sex	6
Ramdan 2018 **[175]**	SEAR	Indonesia	2017-2018	18	19	1-10	N/A	Multivariate logistic regression	Age, location, sex	7
Sein 2016 **[176]**	WPR	Lao PDR	2012-2013	42	79	0-45	Clinical	Bivariate logistic regression	Age, location, sex	8
Vitek 1999 **[177]**	EUR	Russia	1994-1996	58	306	6-8	Laboratory	Conditional logistic regression	Age, class of attendance, location	8
Cohort study										
Chandra 1973 **[178]**	SEAR	India	1971	14	114	0-5	Laboratory	2x2 cross-tabulation	N/A	6
Cross-sectional studies										
Belsey 1969 **[179]**	AMR	United States	1966-1967	249	3,246	N/A	Laboratory	2x2 cross-tabulation	N/A	4
Harnisch 1989 **[180]**	AMR	United States	1974-1975	7	888	N/A	Laboratory	2x2 cross-tabulation	N/A	8
Kalapothaki 1984 **[181]**	EUR	Greece	1980	28	790	6-12	Laboratory	2x2 cross-tabulation	N/A	7
Kitamura 2023 **[182]**	WPR	Vietnam	2019	27	1,189	1-55	Laboratory	2x2 cross-tabulation	Age	9
Marcuse 1973 **[183]**	AMR	United States	1970	131	978	N/A	Clinical or laboratory	2x2 cross-tabulation	N/A	8
Miller 1972 **[184]**	AMR	United States	1970	104	202	N/A	Laboratory	2x2 cross-tabulation	N/A	8
Ohuabunwo 2005 **[185]**	EUR	Latvia	2005	24	281	N/A	Laboratory	Logistic regression	N/A	9

Trichopoulos 1972 **[186]**	EUR	Greece	1970-1971	124	83	6-13	Laboratory	2x2 cross-tabulation	N/A	8
Ecological studies										
Coleman 2018 **[187]**	AMR	United States	1907-1923	221,018	0	N/A	N/A	Linear regression	N/A	-
Dureab 2019 **[188]**	EMR	Yemen	2017	1,294	0	N/A	Clinical	Multivariate logistic regression	N/A	-
Ikejezie 2022 **[189]**	AMR	Haiti	2014-2021	392	0	N/A	Clinical or laboratory	Geographically weighted regression	N/A	-
Izza 2015 **[190]**	SEAR	Indonesia	2010-2011	968	0	N/A	N/A	Correlation tests	N/A	-
Podavalenko 2018 [191]	EUR	Ukraine	1985-2012	21,348	0	N/A	N/A	Binary logistic regression	N/A	-
Quesada 1979 **[192]**	AMR	United States	1970	201	0	N/A	N/A	Stepwise logistic regression	N/A	-

N/A: Not available

Twenty-seven studies were written in English, one in Portuguese, and one in Bahasa Indonesia. Sample sizes (cases and controls combined) ranged from 37 to 221,018 (median=364). Nine studies involved only children (age range=0–14 years); one study only involved adults (age range=40-49 years); six studies involved both children and adults; and thirteen studies did not indicate the age range of the sample. Where reported, the median proportion of male participants was 51% (range=30–93%). Infected individuals were diphtheria cases (20), asymptomatic carriers (six), or a combination of the two (three). Diphtheria was ascertained by isolation of *C. diphtheriae* in culture (12), clinical examination (six), and using either one of the two methods (five); six studies did not specify the method of diagnosis.

2.3.2. Quality of studies according to the Newcastle-Ottawa Scale

All non-ecological studies were of moderate to high quality based on the NOS scores (**Appendix 5**). Both case-control and cross-sectional studies achieved high quality when they implemented advanced statistical techniques and accounted for critical confounders. The assessment also highlighted a trend of greater methodological quality in studies conducted in the 1990s and 2000s compared to older studies.

2.3.3. Potential risk factors

Ninety-five potential risk factors were abstracted from the 29 articles (**Appendix 6**). Altogether, 32 factors were associated with diphtheria in at least one study (**Table 2.2**; **Appendix 7**). Of these, five were related to vaccination or contact with cases, five to underlying conditions, 10 to knowledge and behaviour, two to socioeconomic status, and 10 were population-level factors.

Table 2.2. Factors investigated in the systematic review for which at least one study found an association with diphtheria

Theme and risk factor	No.	Meta-analysis	Risk estimate (95% CI or p value), I^2 Ref.	GRADE score
Vaccination or contact with cases				
Incomplete vaccination (<3 doses)	18	Yes	POR=2.2 (1.4–3.4), I^2=77% [165-168, 170-176, 179, 181-184, 186] N/A [185]	Moderate
No booster vaccination in last five years	4	Yes	POR=3.6 (0.5–24.0), I^2=67% [167, 177] N/A [172, 185]	Very low
Partially vaccinated sibling	1	No	OR=4.1 (1.8–9.0) [169]	Very low
Contact with a diphtheria case	5	Yes	POR=5.0 (0.8–31.8), I^2=75% [171, 173-175] N/A [172]	Low
Contact with a person with skin lesions	3	Yes	POR=4.8 (2.1–10.9), I^2=0% [171, 174] N/A [172]	Low
Underlying conditions				
Having tonsils	2	Yes	POR=2.0 (0.4–10.0), I^2=83% [174, 186]	Very low
Recent sore throat	2	Yes	POR=1.8 (0.8–4.0), I^2=71% [174, 186]	Very low
History of chronic illness	2	No	OR=2.1 (1.2–3.8) [174] N/A [172]	Very low
History of eczema	1	No	OR=3.4 (1.2–9.9) [174]	Very low
Recent fever with myalgia	1	No	OR=2.7 (1.3–5.5) [174]	Very low
Knowledge and behavior				
Sharing utensils, cups, glasses	4	Yes	POR=1.7 (1.0–2.9), I^2=62% [173, 174, 185] N/A [172]	Very low
Sharing a bed or bedroom	4	No	POR=1.3 (0.6–3.0), I^2=76% [173, 174, 176, 182]	Very low
Low diphtheria knowledge	3	Yes	POR=2.4 (1.2–4.7), I^2=20% [164, 170, 175]	Low
Bathing once a day or less	2	No	OR=1.7 (1.04–2.9) [172] p=0.40 [182]	Very low
Bathing once a week or less	1	No	OR=2.6 (1.3–5.2) [174]	Very low
Alcohol abuse	2	No	OR=48.8 (27.2–87.6) [180] OR=0.7 (0.3–1.7) [174]	Very low
Travel history to area with diphtheria	2	Yes	POR=2.2 (0.1–34.1), I^2=84% [173, 175]	Very low
Belief that vaccines are ineffective	1	No	OR=4.0 (1.2–13.5) [164]	Very low

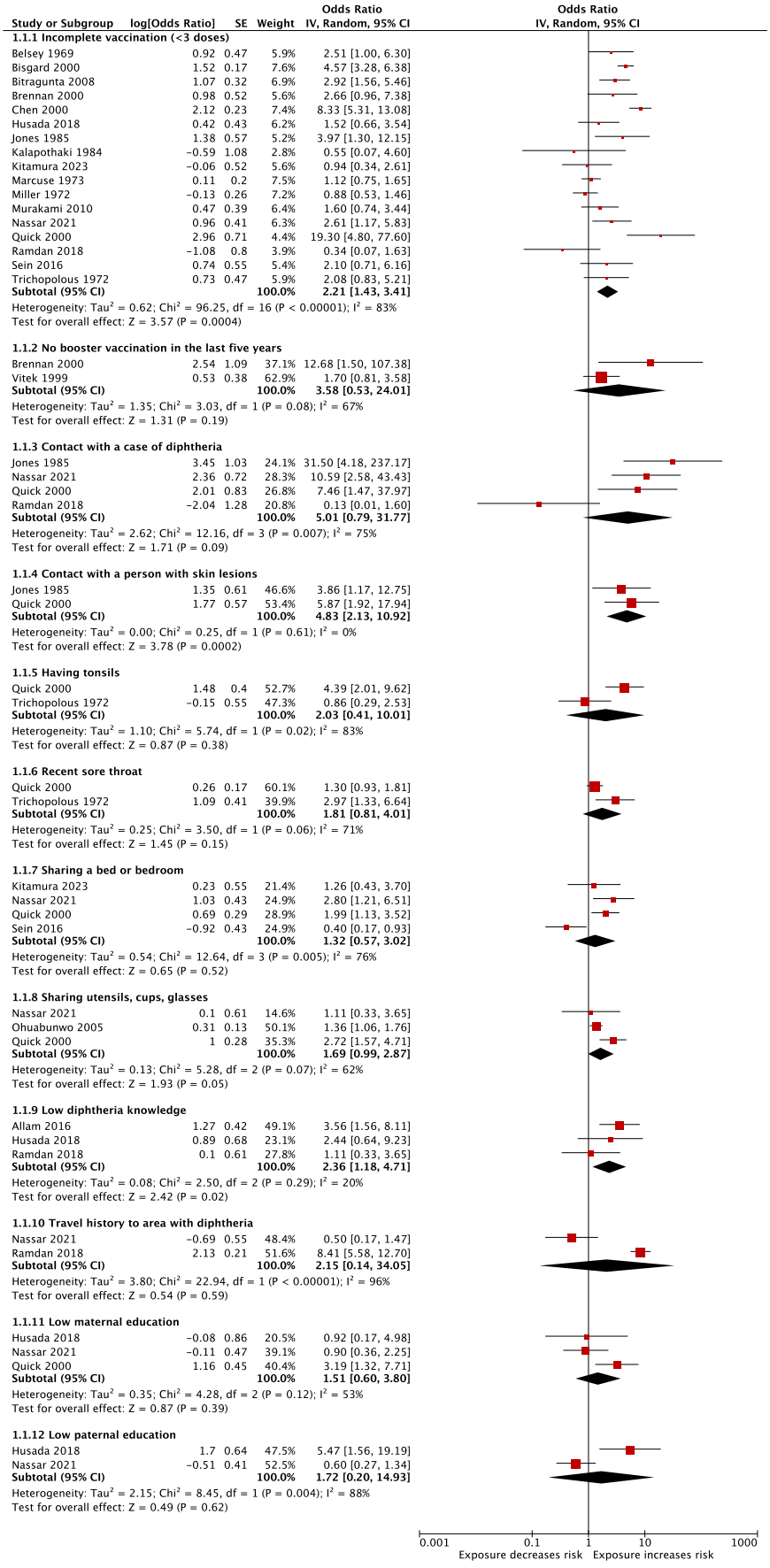
Consumption of factory-made yoghurt	1	No	OR=14.9 (p=0.003) [171]	Very low
Obtaining water from a wheeled carrier	1	No	OR=28.4 (p=0.008) [171]	Very low
Socioeconomic status				
Low maternal education	3	Yes	POR=1.5 (0.6–3.8), I^2=53% [170, 173, 174]	Low
Low paternal education	2	No	POR=1.7 (0.2–14.9), I^2=88% [170, 173]	Low
Population-level factors				
Population density	4	No	ß=0.04 (p<0.001) [191] ß=0.004 (p>0.05) [192] r=0.002–0.07 (p>0.05) [190] ß= -0.001 (p>0.05) [189]	Very low
Vaccination coverage	4	No	ß= -0.04 (p=0.001) [191] r=0.42 (p=0.008); r=0.22 (p=0.183) [190] OR=1.04 (1.01–1.06) [188] ß= 0.177 (p>0.05) [189]	Very low
Degree of urbanization	2	No	ß=0.006 (p<0.01) [189] N/A [191]	Very low
Female literacy	1	No	ß= -0.024 (p<0.001) [189]	Very low
Morbidity rate in the urban population	1	No	ß=0.15 (p=0.001) [191]	Very low
Ongoing conflict	1	No	OR=11.2 (1.3–97.7) [188]	Very low
Percentage of people below poverty line	1	No	ß=0.02 (p<0.05) [192]	Very low
Population growth rate	1	No	ß= -0.23 (p<0.001) [191]	Very low
Sulfur dioxide air levels	1	No	ß=0.23 (p<0.001) [191]	Very low
Tuberculosis cases	1	No	ß=0.03–0.14 (p<0.05); ß=0.03 (p =0.41) [187]	Very low

ß, beta coefficient; OR, odds ratio; POR, pooled odds ratio; r, correlation coefficient.
N/A, effect estimate not available.

Vaccination or contact with cases

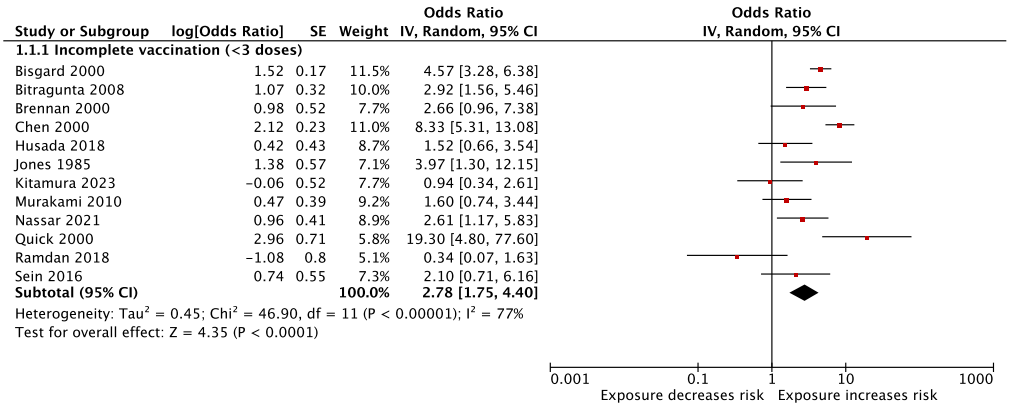
Meta-analysis of 17 studies [165-168, 170-176, 179, 181-184, 186] showed that incomplete vaccination (i.e., having received less than three primary doses of the diphtheria vaccine) doubled the odds of diphtheria (POR=2.2, 95% CI=1.4–3.4) (**Figure 2.2**).

Figure 2.2. Forest plot analysis of potential diphtheria risk factors



A sensitivity analysis was conducted after removing studies that did not adjust for potential confounders (**Figure 2.3**). The analysis was performed only for incomplete vaccination (i.e., having received less than three primary doses of the diphtheria vaccine). The results of the sensitivity analysis (POR=2.8, 95% CI=1.8–4.4) were concordant with the results of the main meta-analysis, in terms of both magnitude and direction of effect.

Figure 2.3. Sensitivity analysis of incomplete vaccination after removing studies that did not adjust for potential confounders.

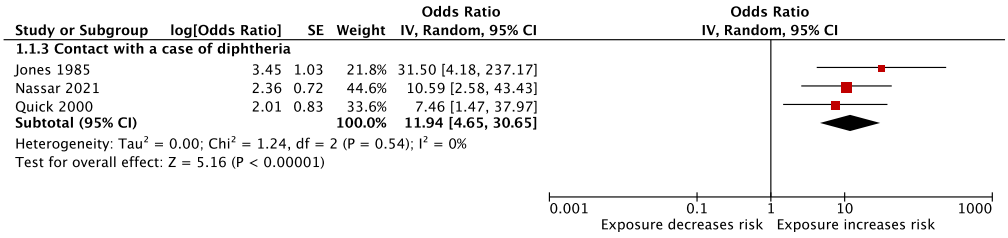


Meta-analysis of two studies conducted in Russia in the mid-1990s [167, 177] found no increased risk of diphtheria associated with not having received a booster vaccination in the preceding five years (POR=3.6, 95% CI=0.5–24.0). Nevertheless, the point estimate was relatively high and the studies had limited power. Two other studies [172, 185] also reported no increased risk of diphtheria associated with time since last booster vaccination. However, these studies did not present effect estimates, nor did they specify the intervals used for the estimations [172, 185]. Finally, data from a study conducted in Brazil in 1969 pointed towards an

increased risk of diphtheria associated with having a partially vaccinated sibling (OR=4.1, 95% CI=1.8–9.0) [169].

Meta-analysis of four studies – two from Yemen [173, 193], one from Georgia [174], one from Indonesia [175] – revealed no increased risk of acquiring diphtheria following contact with a diphtheria case (POR=5.0, 95% CI=0.8–31.8). However, the pooled estimate appeared to be skewed by one study – a preprint article, which also reported outlier values for other risk factors (e.g., incomplete vaccination, low diphtheria knowledge). After removing this study from the meta-analysis, contact with a diphtheria case was associated with an increased risk of diphtheria (POR=11.9, 95% CI=4.7–30.7) (**Figure 2.4**).

Figure 2.4. Sensitivity analysis of contact with a diphtheria case after removing a study that had outlier values for several risk factors.



Meta-analysis of two [171, 174] of the four studies suggest that contact with a person with skin lesions (POR=4.8, 95% CI=2.1–10.9) does increase the risk of diphtheria. A fifth study from Vietnam [172] found no increased risk of infection related to the two exposures. Nevertheless, the study was excluded from the meta-analysis as no effect estimates or data to calculate these were provided.

Underlying conditions

Meta-analysis of two studies from Georgia [174] and Greece (which did not adjust for potential confounders) [186] found that neither having tonsils (POR=2.0, 95% CI=0.4–10.0) or having a recent sore throat (POR=1.8, 95% CI=0.8–4.0) increased the risk of diphtheria. The study from Georgia also reported a doubling of the odds of infection related to having a chronic illness (OR=2.1, 95% CI=1.3–5.5),[174] while a study from Vietnam found no association with the exposure [172]. Nonetheless, the latter study did not provide effect estimates or data to calculate these. Additionally, in the study from Georgia [174], both history of eczema (OR=3.4, 95% CI=1.2–9.9) and recent fever with myalgia (OR=2.7, 95% CI=1.3–5.5) were associated with diphtheria.

Knowledge and behaviour

Meta-analysis of three recent studies from India [164] and Indonesia [170, 175] showed that the odds of disease more than doubled for people with low knowledge of diphtheria (as assessed by a questionnaire) (POR=2.4, 95% CI=1.2–4.7). The study from India also reported an increased risk of diphtheria associated with believing that vaccines do not prevent diseases (OR=4.0, 95% CI=1.2–13.5) [164].

Meta-analysis of three studies from Georgia [174], Latvia (which did not adjust for potential confounders) [185], and Yemen [173] did not identify an increased risk of diphtheria associated with sharing utensils, cups, or glasses (POR=1.7, 95% CI=1.0–2.9). A fourth study from Vietnam [172] found no increased risk of infection related to this exposure. Nevertheless, the study was excluded from the meta-analysis as it did not provide effect estimates or data to calculate these. Meta-analysis of four studies from Georgia [174], the Lao People's Democratic Republic

[176], and Yemen [173] also found that sharing a bed or bedroom with other people did not increase the risk of diphtheria (POR=1.3, 95% CI=0.5–3.9).

Results relating to infrequent bathing were discordant. Two studies from Vietnam examined the association between diphtheria and bathing once a day or less (as opposed to twice a day or more). While one study found no association between the two variables (p=0.40) [182], the other found that bathing once a day or less increased the odds of infection (OR=1.7, 95% CI=1.04–2.9) [172]. Furthermore, a study from Georgia found that bathing less than once a week also increased the risk of diphtheria (OR=2.6, 95% CI=1.3–5.2) [174].

Meta-analysis of two studies from Indonesia [175] and Yemen [173] found no increased risk of infection associated with travel to an area with diphtheria cases (POR=1.3, 95% CI=0.5–3.9). A study conducted in Yemen in the 1980s reported an increase in the odds of diphtheria associated with obtaining water from a wheeled carrier (i.e., a person who carries water obtained from wells to people's houses) (OR=28.4, p=0.008) and consumption of factory-made yoghurt (OR=14.9, p=0.003) [171]. However, the association with yoghurt consumption is likely spurious and may reflect confounding factors or other biases rather than a true causal link.

The study from Georgia also found that having ≥14 alcoholic drinks per week did not increase the risk of diphtheria (OR=0.7, 95% CI=0.3–1.7) [174]. By contrast, data from a study done in the 1970s in the United States (which did not adjust for potential confounders) indicated a

higher risk of disease for "heavy consumers of alcoholic beverages" (OR=48.8, 95% CI=27.2–87.6) [180].

Socioeconomic status

The only socioeconomic indicators that showed a potential association with diphtheria were related to the education level of parents. Meta-analysis of three studies from Indonesia [170], Georgia [174], and Yemen [173] suggested no increased odds of diphtheria associated with having a mother with primary education or less (POR=1.5, 95% CI=0.6–3.8). Similarly, meta-analysis of two studies from Indonesia [170] and Yemen [173] found no increased risk of diphtheria for children whose fathers had primary education or less (POR=1.7, 95% CI=0.2–14.9).

Population-level factors

The above-mentioned risk factors were assessed at the individual level. By contrast, the following factors were examined at the population level in ecological studies comparing reported diphtheria case rates or counts across different areas of the same country.

Four studies from Haiti [189], Indonesia [190], Ukraine [191], and Yemen [188] examined the association between diphtheria and vaccination coverage. The study from Indonesia identified a positive correlation between vaccination level and case count during the first year of a diphtheria outbreak (r=0.42, p=0.008). However, the following year, the two variables were no longer correlated (r=0.22, p=0.183). The study from Ukraine found a negative association between vaccination coverage and the reported case rate in six oblasts (ß=-0.04, p=0.001). This result indicated that a unit increase in vaccination coverage was associated with a 4% average decrease in the case rate, all else being

equal. The study from Yemen showed that vaccination coverage affected the odds of a diphtheria outbreak (OR=1.04; 95% CI =1.01–1.06, p=0.002). In contrast with other articles, the study from Haiti did not find an association between the reported diphtheria case rate and vaccination coverage (ß= 0.177; p>0.05). The authors of the study explained that weaknesses in the country's immunization information systems and inaccuracies in the vaccination coverage estimates may have contributed to the observed lack of association between the two variables.

The association between diphtheria and population density was examined by four studies from Haiti [189], Indonesia [190], Ukraine [191], and the United States [192]. Only the study from Ukraine found a positive association between the reported diphtheria case rate and population density (ß=0.04, p<0.001) [191], whereas the other studies reported no association between the two variables.

A study that analyzed historical records from the early 20th century in the United States found a positive association between diphtheria case counts and reports of tuberculosis (TB) in four of five examined cities (ß=0.06–0.14, p<0.05) [187]. No association was observed in the other city (ß=0.03, p=0.41)

Seven population-level factors were associated with diphtheria in single studies from Haiti [189], Ukraine [191], the United States [192], and Yemen [188]: female literacy (ß= -0.02, p<0.001) [189], morbidity rate in the urban population (ß=0.15, p=0.001) [191], population growth rate (ß= -0.23, p<0.001) [191], presence of an ongoing conflict (OR=11.2, 95% CI=1.3–97.7) [188], sulfur dioxide air levels (ß=0.23, p<0.001) [191], and percentage of people below the poverty line (ß=0.02, p<0.05) [192].

2.3.4. Quality of the evidence based on the GRADE approach

Based on the GRADE assessment, the quality of evidence for the 32 identified risk factors varied from moderate to very low. Evidence for all risk factors was initially rated of low quality due to the observational nature of the included studies. Only the evidence for incomplete vaccination was upgraded from low to moderate given the strength of association with diphtheria observed in the meta-analysis and sensitivity analysis.

Evidence for the other risk factors was often downgraded from low to very low due to concerns related to imprecision as most of the estimates came from a few studies, which in several instances had small sample sizes. Evidence for numerous risk factors was also downgraded due to inconsistencies in the direction of the effect across studies. The risk of methodological bias was considered serious only for population-level factors, whose evidence came from ecological studies. No risk of publication bias was detected for the 32 risk factors. **Appendix 6** illustrates the full assessment of the quality of evidence.

2.4. Discussion

2.4.1. Main findings

This systematic review revealed several factors associated with an increased risk of diphtheria: incomplete vaccination, contact with a person with skin lesions, and low knowledge of diphtheria. Other factors showed potential associations with the disease in multiple studies, including no booster vaccination in the last five years; contact with a case of diphtheria; sharing a bed or bedroom; sharing utensils, cups, and glasses; and infrequent bathing.

The quality of evidence for the identified risk factors was rated as low or very low – except for incomplete vaccination, whose evidence was judged of moderate quality.

Such a finding is not uncommon. Systematic reviews following Cochrane or WHO guidelines are often based on observational evidence rated as low quality [194, 195]. Despite their intrinsic biases and limitations (e.g., confounding), observational studies are essential for public health decision-making – particularly in situations where randomized controlled trials (RCTs) are unavailable, unfeasible, or unethical. In research, ethical considerations generally prevent the deliberate exposure of people to harm, often making RCTs impractical and resulting in a reliance on observational studies for empirical data [196, 197]. Observational studies can provide real-world evidence of the long-term and population-wide effects of health interventions, which RCTs may not comprehensively capture due to their controlled conditions and shorter study periods. The COVID-19 crisis exemplifies this clearly, with observational studies being critical for understanding the spread of SARS-CoV-2, informing public health strategies, and assessing the impact of implemented responses [198, 199].

While GRADE offers a viable approach for evaluating the quality of evidence, it may not be completely suitable for assessing the quality of evidence relating to risk factors like those identified in this systematic review. This is because evidence for this kind of risk factors generally comes from observational studies, whose evidence is always initially rated as low-quality using GRADE – irrespective of the study design or the methodological quality of the studies [149, 150]. When the entirety of the evidence comes from observational studies, it may be advisable to

not consider this as a criterion for quality assessment as this could undermine and conceal the valuable insights provided by this type of research [200, 201]. Instead, the GRADE framework should be adapted, by placing further emphasis on the methodological soundness and the consistency of results across studies. Adopting this approach would lead to a more accurate evaluation of the strength of the evidence and enhance the reliability of findings for policy and practice.

In the sensitivity analysis, incomplete vaccination remained associated with an increased risk of diphtheria after removing studies that did not adjust for potential confounders. This finding was consistent with studies of other vaccine-preventable diseases [202-205], which showed a higher risk of infection for partially vaccinated individuals. This finding was also in line with results of another review [206], which reported that two primary vaccine doses produced lower protective levels of diphtheria antibodies compared with three doses. The importance of vaccines was further corroborated by other studies included in the present review [191, 207], which pointed towards an association between diphtheria and vaccination coverage.

In the meta-analysis, the POR for not having received a booster vaccination in the last five years suggested a potential increased risk of diphtheria. The result may have been inconclusive due to low power and heterogeneity of the study populations. Five years may have also been insufficient to identify time since the last booster as a risk factor, especially in children. Among adults, for which there were fewer studies, this may be a strong risk factor, as suggested by one of the included papers [167]. Previous studies [208-211] found that immunity to diphtheria decreases as age increases. Further research could provide

valuable evidence to inform vaccination guidelines for adults, which currently vary extensively worldwide [212, 213].

Exposure to an infected person is an established risk factor for various vaccine-preventable diseases [202, 204, 214-217]. Hence, it was expected that individuals who had contact with a diphtheria case or a person with skin lesions (a potential proxy for cutaneous diphtheria) would have a higher risk of infection compared to those who did not. These results support current surveillance guidelines that recommend the prompt identification, monitoring, and implementation of preventive measures for close contacts of diphtheria cases [60, 62, 99].

The association between diphtheria and low knowledge of the disease was unsurprising. The increased risk of diphtheria associated with believing that vaccines do not prevent diseases, as reported by one of the included studies, was also expected. Past studies have shown that knowledge of vaccine-preventable diseases and attitudes towards vaccination influence vaccine uptake and adoption of other protective measures [218-220]. It is unclear whether antivaccination attitudes have been increasing in countries that are reporting diphtheria cases. Nonetheless, the recent rise of the internet and social media has certainly created unprecedented opportunities for antivaccination messages and false health information to spread virally, potentially shaping people's knowledge, attitudes, and actions [221, 222].

The increased risk of diphtheria associated with sharing utensils, cups, and glasses; bathing infrequently; and obtaining water from a wheeled carrier (a possible indicator of a lack of access to clean water and sanitation) suggested that poor hygiene may play a role in contracting

the disease. While the exact mechanism through which such practices heighten the likelihood of disease is not fully understood, it is conceivable that they may increase susceptibility to *C. diphtheriae* colonization [174]. The results of this systematic review support those of previous reviews, which concluded that adherence to good hygiene practices lowers the risk of respiratory infections [223, 224].

Incomplete vaccination has often been associated with low parental education [225, 226]. This review also revealed studies indicating a potential relation between education and diphtheria, with children whose parents had low educational levels appearing to be at an increased risk of infection. The pathway behind this association is unclear. Less educated parents may have lower literacy skills that makes them less receptive to health information [227-229]. Less educated parents may also have lower communication abilities, which decrease their capacity to navigate the healthcare system to have their children vaccinated [227, 228, 230]. Finally, less educated parents may simply live in poorer areas where the risk of acquiring diphtheria is higher due to reduced access to healthcare, lower vaccination coverage, and worse sanitary conditions [227, 231, 232]. This last hypothesis seems to be supported by the link identified in one of the studies included in this review between the share of people living below the poverty line and the reported diphtheria case rate.

2.4.2. Implications

Given the strength of association between diphtheria and incomplete vaccination, alongside the consistency of findings from this research with the existing body of knowledge on vaccine-preventable diseases,

additional observational studies on this risk factor would have limited value – except for better characterizing its association with diphtheria. Further research on this subject should, therefore, focus on other factors presenting a possible link with diphtheria and on factors overlooked by previous studies. Overlooked factors include those related to healthcare (e.g., availability of health workers, proximity to health centres, possession of health insurance), which have shown associations with vaccination status [226, 233, 234]. Finally, the risk factors presented in this paper could be used to adjust for confounding in future studies that investigate potential risk factors for diphtheria.

By identifying the main diphtheria risk factors, this review provided a basis for detecting those most vulnerable to the disease and an agenda for public health action. The observed high risk of infection following exposure to a person with skin lesions, alongside evidence in some of the included studies of an increased risk of infection after contact with a diphtheria case, underscored the importance of early case finding and contact tracing. These activities entail the existence of adequate means for laboratory testing. Recent studies highlighted major gaps in diphtheria diagnostics in Europe and in the Western Pacific, including a lack of capacity for molecular typing, insufficient testing equipment, and inadequate training of laboratory personnel [76, 77]. These challenges are likely present in other parts of the world. Tackling them is crucial for controlling diphtheria.

The overwhelming evidence of a link between diphtheria and incomplete vaccination reinforced the need for countries to achieve timely vaccination with a complete primary series followed by booster doses, as recommended by WHO [58]. In the 1990s, countries of the

former Soviet Union succeeded in controlling widespread diphtheria epidemics by vaccinating more than 140 million adults and adolescents alongside millions of children [235]. Since such mass interventions may be difficult for some of the countries currently affected by diphtheria, other solutions should be explored. The recent experience of many LMICs (some of which are reporting high diphtheria rates; e.g., Haiti and India) with delivering the HPV vaccine in schools showed that large numbers of children can be reached through school-based vaccination [106, 107].

Efforts should be made to increase people's knowledge of diphtheria and improve understanding of how to prevent the disease. As healthcare professionals are generally considered the most trusted source of health information [236-239], they should be encouraged to inform patients about diphtheria and the importance of personal hygiene and vaccines. Furthermore, activities should be directed towards implementing effective social mobilization strategies. Polio eradication initiatives in Afghanistan, India, Nigeria, and Pakistan have demonstrated that community health workers can help raise public awareness about vaccination and support surveillance activities [240-243].

This systematic review identified a lack of studies from Africa, despite some countries in this region having reported high rates of diphtheria in recent years. The review also highlighted a paucity of research on diphtheria in Haiti, with only one study from the Haitian context meeting the criteria for inclusion. This may reflect the fact that diphtheria is a relatively rare disease that mainly affects LMICs, where research capacity is sometimes limited [244, 245]. Furthermore, previous

studies have shown that the fees that journals charge to publish articles open access can also pose a barrier for researchers and research institutions in LMICs [246, 247]. The dearth of studies on diphtheria may also be partly due to disparities in research priorities and funding. Røttingen et al. [248] estimated that only 1% of all health research and development (R&D) investments are directed towards neglected diseases like diphtheria, with the majority focusing on ailments that have a significant prevalence and impact in HICs (e.g., cancer, cardiovascular illnesses, Alzheimer's disease and other neurodegenerative disorders). Governments and donor agencies should allocate adequate funding for the implementation of research and dissemination of findings on diphtheria and other vaccine-preventable diseases that disproportionately affect LMICs to strengthen the evidence base and inform disease control efforts. Such research is needed more than ever given the impact of the COVID-19 pandemic, which has caused the largest global disruption to routine immunization services in recent history, leaving millions of children unvaccinated or under-vaccinated against diphtheria – including in countries that had previously controlled the disease [249, 250].

2.4.3. Limitations and strengths

Despite adhering the PRISMA guidelines, this systematic review had several limitations, which should be considered when interpreting the presented findings. Only a single reviewer performed the initial screening of all titles, with relevant studies possibly being missed due to personal interpretation. Although abstracts and full texts were later screened independently by two reviewers, helping to mitigate bias, the initial screening by a single reviewer may have affected the

comprehensiveness of the included studies. Data extraction was conducted manually, increasing the risk of human oversight. While resolving discrepancies through consensus or discussion with a third reviewer is typical for quality assessments using the NOS and GRADE criteria, such approach required subjective decisions that may have influenced the consistency of the results. Identified risk factors emerged from a limited number of single studies, resulting in evidence graded mostly of low quality. The small sample sizes of many of the included studies also meant that studies were not adequately powered to reach definitive conclusions. All studies were observational; as such, they could not demonstrate causality. Certain studies did not report effect estimates for some of the investigated risk factors or data to calculate these. Therefore, it was not always possible to conduct a meta-analysis. Relying on two-by-two table calculations when effect estimates were not available inevitably limited the accuracy of the results. Due to the paucity of studies, it was not possible to stratify the analysis by potential confounders (e.g., age, socioeconomic status). Included studies often varied substantially in methodology. For instance, only a minority of studies (46%) always used laboratory tests to diagnose diphtheria. Thus, other studies may have included individuals who did not contract the infection, possibly leading to an underestimation of the risk estimates. Differences in the methods used for the ascertainment of the exposures, such as surveys versus the review of medical records, may have also affected the estimates. Despite being a cost-effective method for reaching vast populations, surveys are vulnerable to misclassification, recall bias, and social desirability bias due to their self-reporting nature, which might have resulted in imprecise exposure estimates [251, 252]. In contrast, examining medical records can provide more objective data, but challenges might still have arisen relating to the accuracy and

completeness of the records, alongside inconsistent documentation practices [253, 254]. Moreover, half of the studies did not adjust for confounders, which may have contributed to the differences observed across studies on the same risk factors. This methodological heterogeneity may partly explain the high I^2 for some of the pooled estimates.

The high I^2 may also be partly explained by the inclusion of studies from various contexts. The use of an extensive search strategy with clearly defined search terms and no date, geographical, or language restrictions allowed the identification and inclusion of studies from several countries spanning almost all continents, encompassing different time periods and stages of diphtheria epidemics, and covering thousands of individuals across different populations. Such diversity provided more validity to the results of this systematic review and enhanced the generalizability of findings. The employed search strategy helped to minimize reporting bias in this systematic review, although publication bias may still exist in the literature. Furthermore, as with every systematic review, there is a possibility that not all relevant studies were included. All included non-ecological studies were of moderate-to-high quality. The observed higher methodological quality of studies from the 1990s and 2000s compared to older investigations is probably due in part to progress in research practices and improvements in the control of confounding factors over the years. The assessment of study characteristics was conducted meticulously to ensure that only papers with similar exposure definitions would be combined in a meta-analysis. As such, this review did not only provide a comprehensive synthesis of evidence on risk factors for diphtheria, but also offered summary estimates of the association between these exposures and the

disease. These findings are critical since evidence-based policy and practice rely on objective measures of risk, which until now were fragmented and incomplete for diphtheria.

2.4.4. Conclusions

This study identified several risk factors for diphtheria. Moderate to low quality evidence suggested that incomplete vaccination, contact with a person with skin lesions, and low knowledge of diphtheria increased the likelihood of disease. Contact with a case of diphtheria; sharing a bed or bedroom; sharing utensils, cups, and glasses; infrequent bathing; and low parental education also appeared to be associated with diphtheria in multiple studies. Future research should focus on risk factors that have previously been neglected or for which evidence is inconclusive.

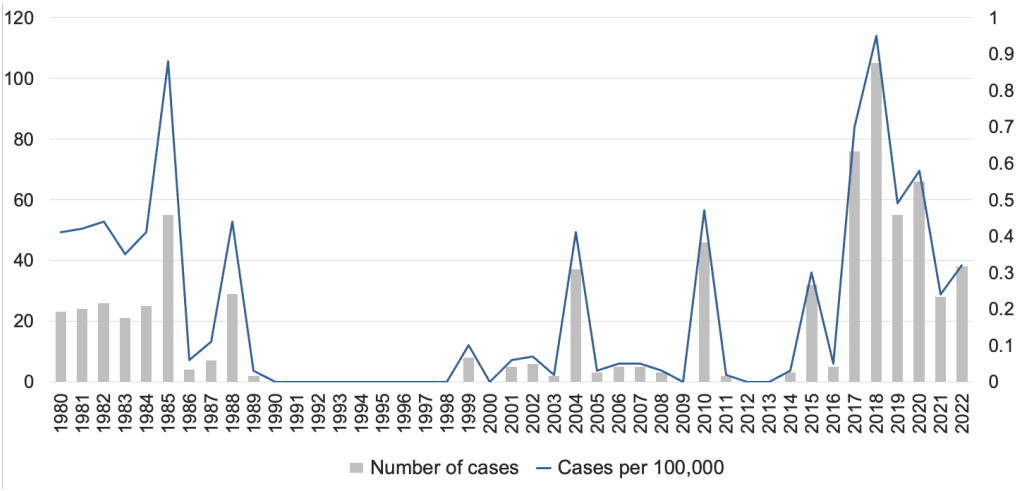
Many factors identified in this review are difficult to modify. Achieving substantial reductions in diphtheria case rates will, therefore, require sustained efforts by countries to strengthen their laboratory capacity, improve vaccination coverage, and increase people's knowledge of the disease and prevention methods with support from health workers. While these interventions have been advocated in the past, the current resurgence of diphtheria makes their implementation as critical as ever – particularly in countries like Haiti, which are heavily impacted by diphtheria yet largely under-resourced and under-represented in research studies.

Chapter 3. The epidemiology of diphtheria in Haiti, December 2014 – June 2021: A spatial modeling analysis

3.1. Introduction

In the Americas, Haiti is among the countries most affected by diphtheria. After the establishment of the EPI, there was a significant decline in reported diphtheria cases, from 216 confirmed cases in the 1980s to eight in the 1990s [59] (**Figure 3.1**). Unfortunately, the progress made did not last. As the country entered the 21st century, the number of cases started to increase again, with outbreaks of varying magnitude reported cyclically. The latest of these outbreaks began in December 2014 and is still ongoing, with 1,750 suspected cases of diphtheria reported as of July 2024 [119]. Among them, 470 were classified as confirmed cases, including 96 deaths (CFR=20.4%).

Figure 3.1. Confirmed diphtheria cases and reported case rates (per 100,000 population) in Haiti, 1980–2022



Data source: WHO. Data accessed on 29 February 2024. This figure is based on data obtained from WHO. Data from 1991 to 1994 were not available. The interpretation and presentation of the data are the sole responsibility of the author and do not represent the official position of WHO.

To date, only two studies have been conducted on diphtheria in Haiti [122, 255]. Both studies, which were carried out during the latest outbreak, found that the majority of the deceased cases were either unvaccinated or had an unknown vaccination status (69–95%). Moreover, the two studies reported a high CFR (31–50%) among confirmed cases. Cited factors for the high proportion of deaths were delays in disease detection, late DAT administration, and limited care [122]. These studies provided some insights into the epidemiology of diphtheria in Haiti, but neither of them provided a comprehensive description of the spatial patterns and dynamics of the disease nor offered clear explanations for its persistence.

Over the last two decades, GIS and spatial analysis have emerged as key tools for detecting disease hotspots and identifying factors that influence

disease transmission [16, 17]. GIS are computer information platforms for collecting, managing, and analysing spatial and non-spatial data. Spatial analysis is a system of non-statistical and statistical techniques that integrates geographic information and numerical data to identify spatial associations that would otherwise be difficult to notice.

Few studies have employed GIS and spatial analysis to examine the epidemiology of diphtheria in different settings. For instance, Podavalenko [191] detected a significant correlation between reported diphtheria case rates and vaccination coverage, population density, and population growth rate in Ukraine from 1985 to 2016. Nailul *et al*. [207] also identified a negative association between reported diphtheria case rates and vaccination coverage in East Java, Indonesia in 2010. Another study by Quesada *et al.* [256] revealed significant associations between diphtheria case rates and poverty rates during an outbreak in San Antonio, Texas in 1970.

Despite the growing body of literature in this field of research, there remain significant knowledge gaps regarding the spatial epidemiology of diphtheria, particularly within the context of Haiti. No studies have investigated the spatial patterns of diphtheria transmission in the country. Thus, it is critical to develop explanatory spatial models that can enhance the general understanding of the dynamics and determinants of the disease's occurrence in a resource-limited setting such as Haiti.

This chapter presents a study that leveraged GIS and spatial analysis to address the aforementioned research gaps. The present study set out to characterize the spatial epidemiology of diphtheria in Haiti from

December 2014 to June 2021. Specifically, it aimed to determine the subnational distribution of confirmed cases in the country, locate hotspots of transmission, and identify potential factors associated with the reported case rate of the disease.

The following sections of this chapter describe the research context and illustrate the methods employed in this study. The remainder of this chapter reports the results of the spatial modeling analysis and discusses research findings in the context of the relevant diphtheria literature. The chapter ends by presenting key conclusions of the study, summarizing the main policy and practice implications of the findings, and recommending areas for future research.

3.2. Research context

3.2.1. Study area

Haiti, officially the Republic of Haiti, is a country situated on the western third of Hispaniola (19.00° N latitude, 72.25° W longitude) – an island in the Caribbean Sea that it shares with the Dominican Republic [257, 258]. Haiti is 27,750 km^2 (10,714 square miles), making it the third largest country in the Caribbean by surface area.

Haiti is divided into 10 departments: Artibonite, Centre, Grande Anse, Nippes, Nord, Nord-Ouest, Nord-Est, Ouest, Sud, and Sud-Est [257] (**Figure 3.2**). These departments consist of 42 arrondissements, 140 communes, and 570 communal sections. The capital and largest city is Port-au-Prince, which is in the Ouest department. Over one-third of the population lives in the Ouest Department, where the country's capital is located.

Figure 3.2. Map of Haiti's ten departments

Most of Haiti's territory is mountainous, with about two-thirds of the land area above 490 metres in elevation [118]. The country has a warm, humid tropical climate, with some variation depending on altitude. Average temperatures range from 25 °C in January to 30 °C in July. In winter, frost can occur at high altitude. Haiti has two major seasons: a rainy summer (from May to October) and a dry winter (from November to April) [118].

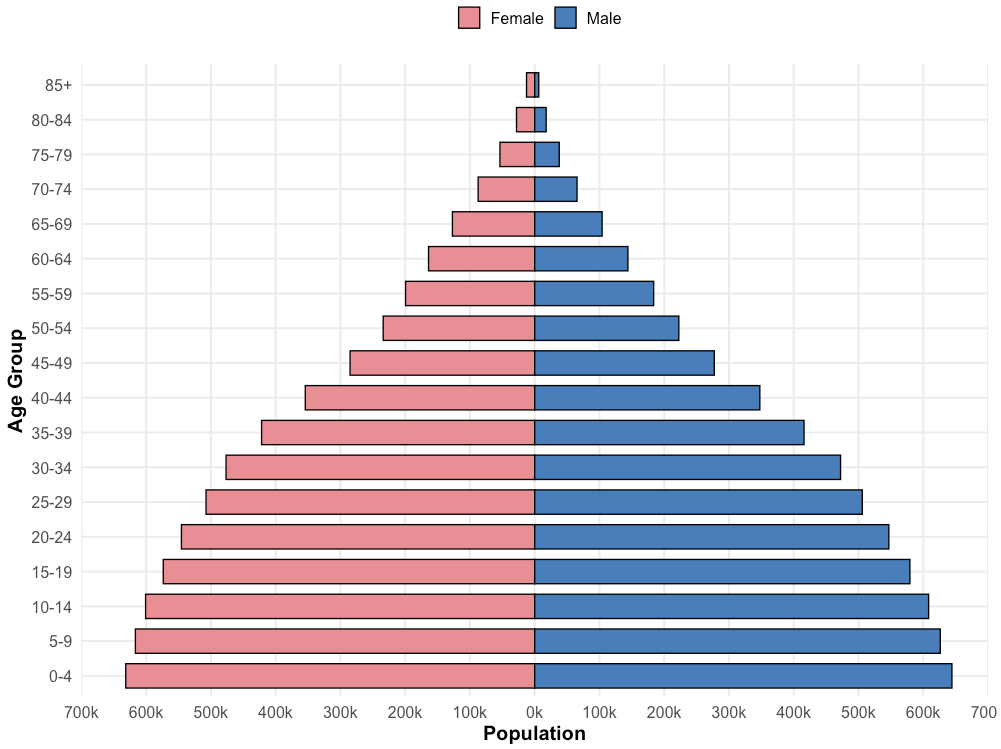
Haiti is situated in the middle of the hurricane belt, which makes the country susceptible to severe storms during the Atlantic hurricane season (from June to October) [259]. In recent history, Hurricane Matthew – a Category 4 hurricane that made landfall on 4 October 2016 – was particularly damaging for Haiti, causing a death toll of 546, displacing 175,000 people, and leaving 1.4 million people in need of humanitarian assistance [260].

The country is also located in an active seismic zone. On 14 August 2021, an earthquake with a 7.2 magnitude on the Richter scale hit southwestern Haiti, leaving at least 2,248 people dead, and more than 12,200 injured [260]. Nevertheless, the most devastating earthquake in recent years hit the country on 12 January 2010, causing more than 300,000 deaths and displacing over 1.5 million people [261].

3.2.2. Demography

Haiti is the most populous Caribbean country, with a population estimated at about 11 million [257]. The population density is high, at an average of 409 people per km^2. The male to female ratio is about even, with 98 males for every 100 females [261] (**Figure 3.3**). The Haitian population is very young, with a median age of 24 years. An estimated 31% of the total population is under 15 years of age. The current life expectancy stands at 66 years. In 2016, the literacy rate among people aged 15 years and above was 62% [261]. Nearly 95% of the population is of African descent, while the remainder is mostly of mixed African and European ancestry [118].

Figure 3.3. Population pyramid of Haiti, 2023



Data source: World Bank. Data accessed on 12 May 2024. This figure is based on World Bank data. Interpretation and presentation of the data are the responsibility of the author and do not represent the official position of the World Bank.

There are two official languages in Haiti: French and Haitian Creole [118]. Although French is the primary language of formal education, it is spoken fluently only by 10% of the population. By contrast, nearly all of the population speaks Creole.

There is no official religion in Haiti. About 55% of the population practices Roman Catholicism, while 29% is Protestant [261]. Around 2% of Haitians also practice Voodoo, a religion with West African roots.

3.2.3. Politics

Haiti is a multi-party parliamentary representative democratic republic [118]. The president is the head of state and is elected by direct popular vote for a term of five years. The prime minister is the head of government, selected by the president and confirmed by the National Assembly [118].

The history of Haiti has been marked by periods of civil unrest triggered by multiple coups, military regimes, and economic crises [118]. However, since 2018, there has been a significant escalation in the level of political turmoil, with recurrent, country-wide protests against successive governments. These protests have been driven by fraud, bribery, and corruption allegations; soaring inflation; and the absence of basic services (e.g., electricity and clean water) for most of the population. In September 2018, the protests led to the resignation of Prime Minister Jack Guy Lafontant and the formation of a new government [262]. Nevertheless, civil unrest continued, reaching new heights in July 2021, when Haiti's President Jovenel Moïse was assassinated by unidentified gunmen in his private residence in Port-au-Prince [118]. Ariel Henry, who was appointed prime minister in the days preceding the assassination, became interim president. However, he has subsequently been accused by the country's head prosecutor of communicating with a key suspect hours after the assassination. Presidential elections have been postponed several times, and have yet to be held as of August 2024 [263].

3.2.4. Economics

As explained above, Haiti is still experiencing political turmoil, rising violence, and unprecedented levels of insecurity. Collectively, these factors – alongside the country's weak infrastructure – are hindering national socioeconomic development. In 2020, Haiti ranked 170 out of 189 countries on the United Nation (UN)'s Human Development Index [264].

In 2021, the country had a gross domestic product (GDP) per capita of US$1,830 – the lowest in the Region of the Americas and less than a quarter of the regional average of US$8,328 [265]. Nearly 60% of the population lives below the national poverty line [266]. The unemployment rate was projected to be at 16% in 2022 [267]. In 2017, it was estimated that the primary sector (i.e., agriculture, fishing, forestry) accounted for 22% of the overall GDP; the secondary sector (i.e., manufacturing, energy production, construction), 20%; and the tertiary sector (i.e., commerce, hospitality), 58% [261].

The national currency is the Haitian gourde (HTG). In recent years, the gourde has depreciated against the US dollar, going from about 55 HTG in July 2015 to approximately 115 HTG in July 2022 [268]. This has contributed to a rise in the cost of living, as most products are imported and paid for in US dollars, requiring increasingly more gourdes for their purchase.

3.2.5. Healthcare

In the 2017-2018 biennium, the share of the government's budget designated for health was 4% – around half the average for LICs [266].

In 2015, per capita health expenditure was about US$32, which is significantly lower than the estimated US$86 required to ensure an essential package of services in LICs [266]. External financial cooperation accounts for 80% of non-private health expenditure [266].

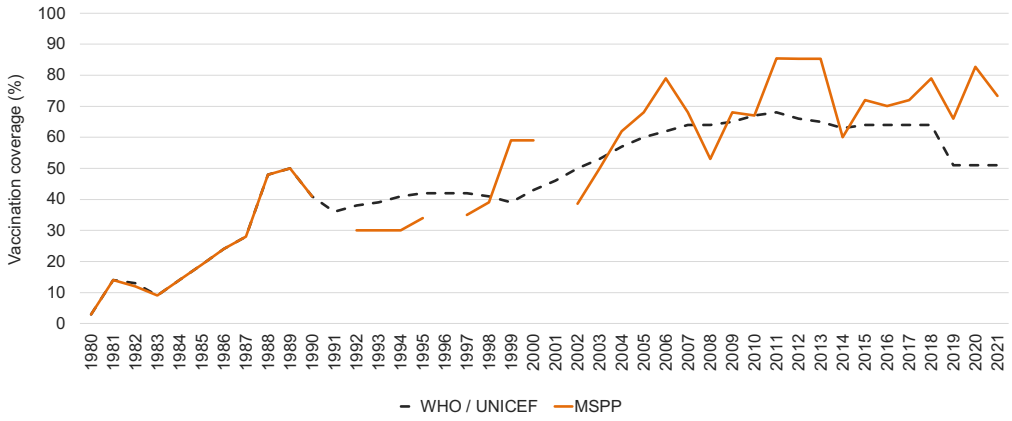
The national healthcare system consists of MSPP; ten Departmental Health Directorates; the Arrondissement Health Units; and healthcare facilities [269]. MSPP is responsible for supervising, evaluating, and monitoring service delivery; policymaking; and the health budget.

In 2015, there were 1,067 healthcare facilities in Haiti [270]. These facilities belong to one of four sectors: public, private nonprofit, private for-profit, and mixed [271]. Roughly 38% of healthcare facilities are in the public sector. The private nonprofit sector, which accounts for 18% of all healthcare facilities, consists of nongovernmental organizations (NGOs). About 24% of all healthcare facilities are part of the private for-profit sector, which comprises physicians, dentists, nurses, and other specialists who work in clinics in Port-au-Prince and other large cities. The remaining 20% of healthcare facilities make up the mixed sector, in which the staff is paid by the State, but management is handled by private entities [271].

Just 23% of people in Haiti live within 5 km of a dispensary (i.e., the main facility for primary care provision) [271]. The country has 0.3 dispensaries per 10,000 population – well below the expected ratio of 1 dispensary per 10,000 population [271]. Access to healthcare centres (i.e., the second level of primary care) is higher and aligned with other LICs, with a ratio of 1.2 health centres per 30,000 population [271]. Yet, these centres are rarely equipped to provide the appropriate level of care.

Haiti's health outcomes are poor, even when compared to other LICs. Haiti fares especially poorly with deliveries at healthcare facilities and immunization coverage, with high inequalities across wealth quintiles. For example, the percentage of mothers who deliver in healthcare facilities in Haiti (37%) is nearly half of what is observed in other countries (70%) [271]. Haitian mothers are also less likely to deliver in a healthcare facility if they are in the lowest household income quintile (9%) compared to the highest (76%) [271]. In 2021, according to the MSPP, 73% of children under 24 months received the primary three doses of the DTP vaccine (DTP3); estimates by the WHO and UNICEF indicate a lower vaccine coverage (51%) [272] (**Figure 3.4**).

Figure 3.4. DTP3 vaccination coverage in Haiti according to MSPP and WHO / UNICEF estimates, 1980 – 2021

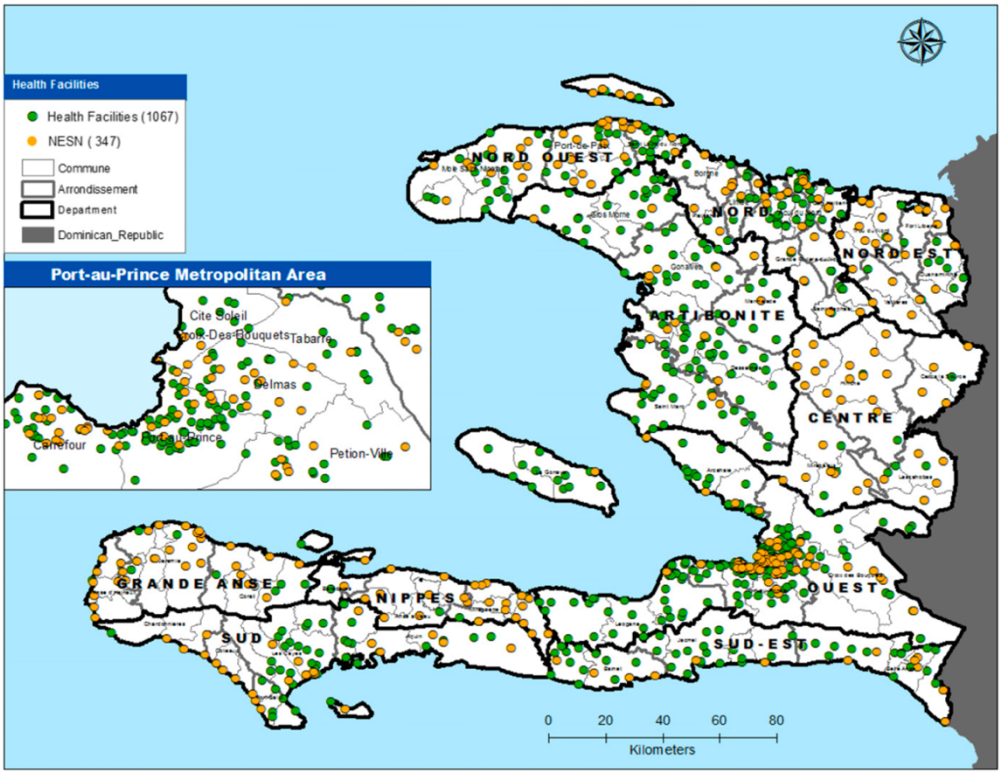


Data source: WHO and MSPP. Data accessed on 24 April 2022. MSPP estimates for 1991, 1996, and 2001 were not available. The interpretation and presentation of the data are the sole responsibility of the author and do not represent the official position of WHO or MSPP.

3.2.6. Surveillance

Since 2005, public health surveillance in Haiti has been the responsibility of the Directorate of Epidemiology, Laboratory and Research (Direction d'épidémiologie, des laboratoires et de la recherche; DELR) – an arm of the MSPP. In 2010, following the earthquake, the DELR launched the National Epidemiologic Surveillance Network (NESN), to help detect and respond to outbreaks [273]. Between 2010 and 2015, the number of sentinel sites increased from 51 to 347 [269] (**Figure 3.5**).

Figure 3.5. Geographic distribution of healthcare facilities and sentinel surveillance sites in Haiti, 2015



Reproduced from Juin *et al*. (2017), licensed under CC BY 4.0. Source: The American Journal of Tropical Medicine and Hygiene, 97(Suppl 4), 12–20 [269].

As part of the NESN, every week, epidemiologic surveillance officers collect data from healthcare facilities on 47 diseases and events – 14 of

which are immediately reportable, including diphtheria [269]. After review and validation, the data are entered into web-based platforms, allowing direct reporting to the departmental health directorates and the DELR [269]. Weekly surveillance meetings are convened at both the departmental and central level to discuss epidemiological analyses and inform action pertaining to ongoing public health events [269].

3.3. Methods

3.3.1. Study design

The study aimed to explore the spatial epidemiology of diphtheria in Haiti, using a retrospective ecological analysis of confirmed cases. In this study, a confirmed case was defined as an individual who tested positive for *C. diphtheriae* by PCR or who was confirmed by epidemiological link, in line with case definitions and surveillance practices adopted by the DELR.

The geographical unit of analysis was the commune, which was selected to obtain granular spatial insights that could potentially inform public health action. The period under consideration was from 1st December 2014 to 30th June 2021, which had been chosen based on data availability. By covering multiple years, this period offered ample breadth to capture long-term trends and patterns of diphtheria, thereby enabling a thorough understanding of the spatial distribution and temporal dynamics of the disease.

3.3.2. Data dictionary

In this study, we used the expression "reported diphtheria case rate" instead of "diphtheria incidence" to avoid implying that the figures

presented here are an exact measure of new infections in a clearly defined population at risk over a specified interval. This terminology better reflects the scope of the analysis and its underlying data limitations.

The number of diphtheria cases at the commune level was obtained from the DELR. Crude annual rates by communes were calculated by dividing the number of diphtheria cases reported annually by the corresponding population estimate from the Haitian Institute of Statistics and Informatics (Institut haïtien de statistique et d'informatique; IHSI) [257]. Average rates were calculated by dividing the sum of the total cases reported during the study period by the sum of the populations for the same period. All rates were multiplied by 100,000.

Eleven factors which could be linked to reported diphtheria case rates were selected following the systematic review presented in Chapter 2 [274]. These were grouped under three domains: health, socioeconomic status, and environment. **Table 3.1** summarizes the study variables. A direct measure of poverty could not be included in the analysis due to a lack of up-to-date, reliable, and spatially disaggregated data for this factor, hindering our ability to explore its potential influence on reported diphtheria case rates.

Table 3.1. Variables selected for the spatial modeling analysis

Theme and variable	Description	Source and study period
Reported diphtheria case rate	Confirmed diphtheria cases per 100,000 population	DELR, 2014–21
Health		
Coverage for the third dose of the diphtheria tetanus pertussis (DTP3) vaccine	Proportion of children aged <1 year who had received the third dose of the DTP vaccine	MSPP, 2015–20
Diphtheria tetanus (DT) vaccine stockout	Average annual number of days when the DT vaccine was out of stock	MSPP, 2017–20
DTP stockout	Average annual number of days when the DTP vaccine was out of stock	MSPP, 2017–20
Health facility density	Number of healthcare facilities per 100,000 population	Humanitarian Data Exchange, 2020
Socioeconomic status		
Female literacy	Proportion of women who are literate	DHS, 2016–17
Improved water source	Proportion of the population that lives in households whose main source of drinking water is an improved source	DHS, 2016–17
Male literacy	Proportion of men who are literate	DHS, 2016–17
No toilet facility	Proportion of the population that lives in households with no toilet facility	DHS, 2016–17
School density	Education facilities per 100,000 population	DHS, 2020
Environment		
Population density	Ratio between total population and total surface area	IHSI, 2015
Urbanization	Proportion of urban population in total population	IHSI, 2015

Data for most of these variables were extracted from spatially interpolated maps produced by the Demographic and Health Survey (DHS) Program [275]. The maps were freely available as raster files on the DHS Program Spatial Data Repository. The maps were based on a 2016–2017 survey of a nationally representative sample of 13,405 households in Haiti [276]. Using a simple mean approach, datapoints in

the maps were aggregated to match the boundaries of each commune using R programming language [277]. Spatial data relative to administrative boundaries and healthcare facilities in Haiti were retrieved from Humanitarian Data Exchange (HDX) – an open access platform managed by the United Nations Office for the Coordination of Humanitarian Affairs (UN OCHA) [278]. Other data sources included the MSPP and the IHSI.

3.3.3. Ethical considerations

Since all datasets used in this study were anonymized and aggregated at the commune level, no consent was required. The study was approved by Haiti's National Bioethics Committee (reference number: 1921-45) (**Appendix 8**) and by the University of Nottingham's School of Medicine Research Ethics Committee (reference number: 267-1903) (**Appendix 9**).

3.3.4. Descriptive analysis

Collected data were examined for consistency by checking for missing, duplicate, and out-of-range values. Frequency distributions were generated for categorical variables. Measures of location (i.e., mean, median) and variation (i.e., standard deviation, range, interquartile range) were calculated for continuous variables. Choropleth maps were developed to illustrate the geographic distribution of the study variables. QGIS [279] was used to process data while the descriptive analysis was performed using R programming language.

Two variables (DT vaccine stockout and DTP vaccine stockout) were excluded from the analysis due to the large amount of missing data

(>10%), which could have affected the accuracy and reliability of the results. Out-of-range values were found for DTP3 vaccine coverage. However, since these values represented less than 10% of the total number of observations, the variable was included in the analysis. No duplicate values were found in the dataset.

3.3.5. Spatial autocorrelation and hotspot analysis

Spatial autocorrelation analyses were conducted to investigate the spatial pattern of reported diphtheria case rates and identify hotspots. The global spatial test Moran's I was used to quantify the spatial autocorrelation of the diphtheria case rate in Haiti. The Moran's I is an index that measures the extent of spatial autocorrelation in a given dataset using a scale from -1 to +1 [280, 281]. A positive Moran's I suggested positive autocorrelation (i.e., the clustering of communes with similar values). A negative Moran's I denoted negative autocorrelation (i.e., the clustering of communes with dissimilar values). A Moran's I close to 0 indicated that values were randomly distributed.

Since the global Moran's I revealed the overall degree and direction of spatial autocorrelation but not where the clustering of high and low values occurred, local indicators of spatial association (LISA) were also calculated. LISA are a local version of the Moran's I, in which the level of spatial clustering is assessed around each individual geographical unit (e.g., commune) rather than across the entire study area (e.g., Haiti) [282]. In this study, neighbour relationships were defined using a first-order Queen's contiguity method, in which only communes that shared common boundaries were considered to be neighbours. If a commune was situated on an island and, thus, did not share borders with the rest

of the study area, these were assigned manually to the nearest commune on mainland Haiti [283]. The main output of the LISA analysis was a map showing four types of statistically significant spatial autocorrelation [282]: high-high, indicating clusters of communes with high reported diphtheria case rates (i.e., the hotspots); low-low, showing clusters of communes with low case rates (i.e., the cold spots); and low-high and high-low, representing spatial outliers (i.e., communes with low case rates surrounded by those with high case rates, and vice versa).

Alternative methods that could have been used for this analysis include Ripley's K function, Kulldorff's spatial scan statistic, and Getis-Ord Gi*. Ripley's K function assesses whether the distribution of points (e.g., cases of a disease) is clustered, dispersed, or random by calculating the number of points within different distances from each other and comparing it to what would be expected under randomness [284]. Kulldorff's scan statistic detects clusters of cases by moving a circular or elliptical window of varying sizes across a given area to determine whether the number of cases occurring within the window is higher than expected under random distribution [285]. The Getis-Ord Gi* statistic identifies clusters by comparing the number of cases within each location (e.g., a district) to the overall distribution of cases across the entire study area (e.g., a country) to indicate where cases occur more or less frequently than predicted by random chance [286].

Ultimately, the Moran's I and LISA were chosen instead of the other methods for multiple reasons. Unlike Ripley’s K function, Moran's I and LISA directly measure spatial autocorrelation, thereby revealing whether locations with high reported case rates are surrounded by other areas of similarly high case rates [287]. LISA are more computationally

efficient than Kulldorff's spatial scan statistic and do not necessitate prior assumptions about the size or shape of clusters, reducing the degree of complexity and making them more suitable for exploratory investigations [287]. While Getis-Ord Gi* also identifies hotspots and cold spots, LISA go a step further by considering low-high and high-low outliers, leading to a more nuanced illustration of local spatial relationships [282].

All spatial analyses were conducted in GeoDa 1.12 [288]. The level of significance was set at $p<0.05$. The significance of the spatial tests was evaluated by comparing the observed test results with the expected results under the complete spatial randomness assumption using Markov chain Monte Carlo method based on 999 permutations [289].

3.3.6. Regression models

To identify the significant correlates of the reported diphtheria case rate, two regression models were built: ordinary least squares (OLS) and geographically weighted regression (GWR). OLS is a global model which presumes that observations are mutually independent and that relations between dependent and independent variables are constant across a study area. When these assumptions are violated, global models are no longer effective. OLS is defined as [290]:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots \beta_n X_n + e \quad (1)$$

where Y is the dependent variable, X is the independent variable, β is the coefficient explaining the strength and type of relationship between X and Y, and e is the residual (i.e., the difference between the observed

value of Y and the value of Y predicted by the model). The residual is a proxy for ϵ, which is the error term (i.e., the difference between the observed value of Y and the unobserved true value of Y). The error term represents all the factors influencing Y that the model does not capture.

In the context of this study, the OLS model was specified as follows:

$$Reported\_Diphtheria\_Case\_Rate = \beta_0 + \beta_1 DTP3\_Coverage_1 + \beta_2 DT\_Stockout_2 + \beta_3 DTP\_Stockout_3 + \beta_4 Health\_Facility\_Density_4 + \beta_5 Female\_Literacy_5 + \beta_6 Improved\_Water\_Source_6 + \beta_7 Male\_Literacy_7 + \beta_8 No\_Toilet\_Facility_8 + \beta_9 School\_Density_9 + \beta_{10} Population\_Density_{10} + \beta_{11} Urbanization_{11} + e \quad (2)$$

In contrast with OLS, GWR is a local model that accounts for spatial heterogeneity by generating a unique equation for every unit of a study area [287, 291]. Each equation is calibrated based on their neighbouring units, which are weighted using a decreasing function of distance; in other words, nearby areas hold a greater weight than those farther away. The assumption is that everything is related to everything else, but near things are more related than distant things (i.e., Tobler's first law of geography) [292]. GWR can be defined as:

$$Y_i = \beta_{0i} + \beta_{1i} X_{1i} + \beta_{2i} X_{2i} + \ldots \beta_{ni} X_{ni} + e_i \quad (3)$$

in which i is the specific location where data on Y and X are measured.

In this analysis, the GWR model was formulated as follows:

$$Reported\_Diphtheria\_Case\_Rate_i = \beta_{0i} + \beta_{1i}DTP3\_Coverage_{1i} + \beta_{2i}DT\_Stockout_{2i} + \beta_{3i}DTP\_Stockout_{3i} + \beta_{4i}Health\_Facility\_Density_{4i} + \beta_{5i}Female\_Literacy_{5i} + \beta_{6i}Improved\_Water\_Source_{6i} + \beta_{7i}Male\_Literacy_{7i} + \beta_{8i}No\_Toilet\_Facility_{8i} + \beta_{9i}School\_Density_{9i} + \beta_{10i}Population\_Density_{10i} + \beta_{11i}Urbanization_{11i} + e_i \quad (4)$$

The final independent variables to be included in the two models were identified using a multi-stage process to ensure the absence of multicollinearity, which occurs when independent variables are highly correlated among each other [293]. Firstly, Spearman's rank correlation was conducted to identify strong correlations ($r \geq 0.7$, $p \leq 0.05$). If two or more independent variables were highly correlated, the one with the lowest correlation with the reported diphtheria case rate was excluded. Then, the remaining variables were included in the OLS model. Finally, the variance inflation factor (VIF) was calculated to determine the degree of multicollinearity among the independent variables. A VIF≤5 was considered acceptable. Variables that did not show a statistically significant ($p > 0.1$) effect on the diphtheria case rate were removed from the model.

The performance of the OLS and GWR models was compared using the adjusted r-squared (R^2) and the corrected Akaike information criterion (AIC_c). R^2 is the coefficient of determination, which indicates the proportion of variance in the dependent variable that is collectively explained by the independent variables [294]. A drawback of R^2 is that it increases with the number of added variables. The adjusted R^2 is similar to the ordinary R^2, but it imposes a penalty as superfluous

variables are included in the model. AIC_c is a modified version of the Akaike information criterion (AIC), a comparative measure of goodness-of-fit that takes into account model complexity [295]. AIC is obtained by the sum of twice the negative log-likelihood and twice the number of parameters in the model. Lower AIC scores are indicative of higher efficiency (i.e., models that explain a greater amount of variation using fewer parameters). AIC_c is equivalent to AIC but with a correction for small sample sizes.

For this analysis, alternative methods were also considered, such as the Spatial Autoregressive Model (SAR), the Spatial Error Model (SEM), and Bayesian Hierarchical Models. SAR is a global model that incorporates spatial autocorrelation by assuming that the value of the dependent variable at one location is influenced by the values at surrounding locations [296, 297]. SEM is another global model that accounts for spatial autocorrelation in the error terms, which implies that unobserved factors influencing the dependent variable are spatially correlated [296, 297]. Bayesian Hierarchical Models account for random variations at different geographical scales (e.g., individual, neighborhood, and region) and incorporate prior information about the model variables to enhance accuracy, providing a robust framework for measuring the uncertainty in parameter estimates and making probabilistic inferences [298, 299].

While GWR captures local variations in relationships, SAR and SEM assume global spatial processes [38, 296], which can lead to spatial heterogeneity being missed. Furthermore, SAR and SEM require the specification of spatial weights matrices and overall estimation techniques that are more complex than those of GWR [38, 296]. Although

Bayesian Hierarchical Models can handle complex spatial structures, they require advanced computation and specialized expertise in Bayesian statistics [298-300], which represented critical barriers given the limited resources and time constraints of this study. By choosing to use OLS and GWR, the aim was to strike a balance among simplicity, interpretability, and computational efficiency. While OLS provided a foundational understanding of the relationships between the reported diphtheria case rate and the independent variables, GWR allowed to explore how these relationships varied across space.

Results output from the GWR model were used to create surface maps of the R^2 values and local coefficients of each independent variable to explore the spatial variation in the relationship between the reported diphtheria case rate and the selected parameters. All regression models and surface maps were developed using R programming language.

3.4. Results

3.4.1. Descriptive analysis

From December 2014 to June 2021, 392 confirmed diphtheria cases were recorded in Haiti (**Table 3.2**). Most of the cases were female (n=215; 54.8%) and aged ≤14 years old (n=343; 87.5%). Only 59 cases (15.1%) were reported to be vaccinated against diphtheria, which was defined as having received at least three doses of a diphtheria vaccine.

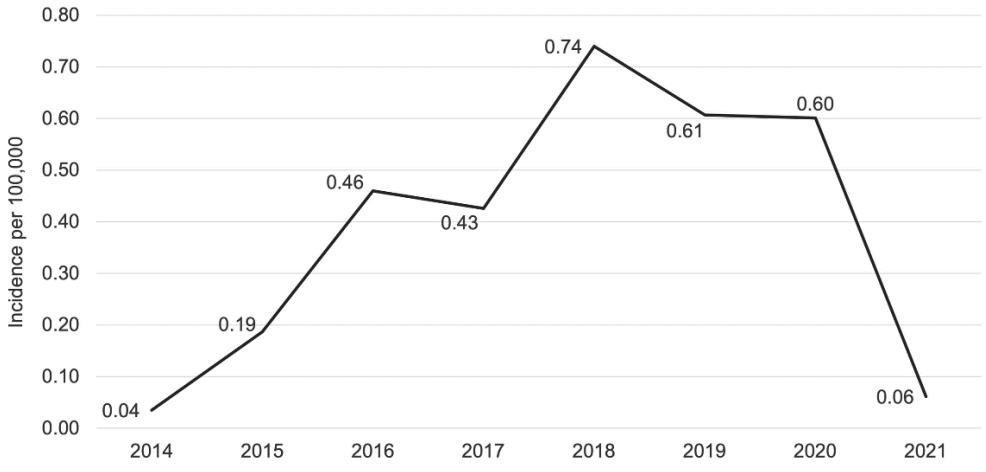
Table 3.2. Characteristics of confirmed diphtheria cases in Haiti, December 2014 – June 2021

Characteristics	n (%)
Total confirmed cases	392
Female	215 (54.8)
Male	177 (45.2)
Age (in years)	
<5	84 (21.4)
5–14	259 (66.1)
>14	49 (12.5)
Vaccination status	
n/a*	4 (1.0)
Unknown	209 (53.3)
Unvaccinated	120 (30.6)
Vaccinated	59 (15.1)

* Cases for which information on the vaccination status was not available.

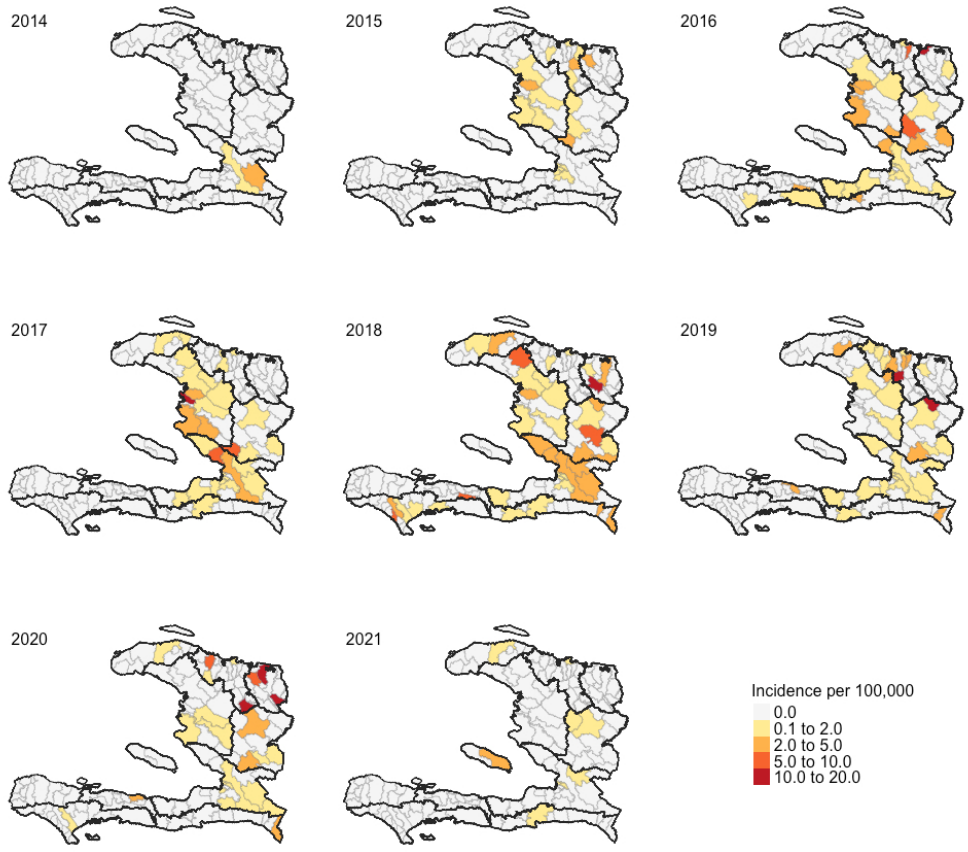
During the study period, the average annual reported diphtheria case rate varied greatly, going from 0.04 cases per 100,000 population in 2014 to 0.74 per 100,000 in 2018 (**Figure 3.6**). This peak was followed by a three-year decline in reported infection rates.

Figure 3.6. Average annual reported diphtheria case rate (per 100,000 population) in Haiti, December 2014 – June 2021



Information on the commune of origin was not available for two of the 392 cases. As **Figure 3.7** shows, the outbreak appeared to originate in the Ouest department and to have gradually spread to the rest of the country. Between 2014 and 2015, detection of diphtheria cases remained limited to 21 communes across five departments located in central and northern Haiti. By 2021, cases had been reported in 79 communes, encompassing nine departments. Grand'Anse was the only department to report no confirmed cases throughout the study period. Four departments (i.e., Artibonite, Centre, Nord, and Ouest) accounted for 84% of all confirmed cases. Ouest was the only department to report cases each year.

Figure 3.7. Reported diphtheria case rate (per 100,000 population) in Haiti, December 2014 – June 2021

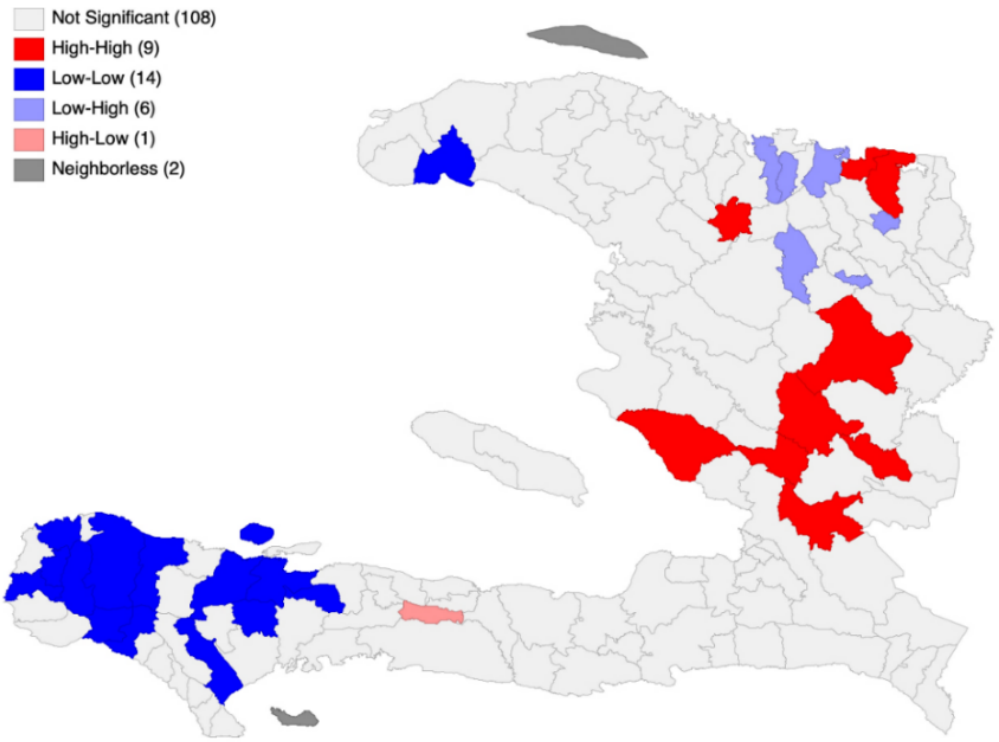


3.4.2. Spatial autocorrelation and hotspot analysis

The global Moran's I test found modest but statistically significant spatial autocorrelation ($I = 0.18$, $p < 0.001$). This suggests that, during the study period, reported diphtheria case rates were more similar in certain neighbouring communes than would be expected by chance.

The LISA analysis revealed nine communes, home to an estimated 646,346 people (4.7% of the population of Haiti), that can be classified as diphtheria hotspots (**Figure 3.8**). Furthermore, one high-low commune (i.e., a commune with a high reported case rate surrounded by communes with low reported case rates) was found in the Sud department. An estimated 35,139 people (0.3% of the population) live in this high-low commune. Additionally, the analysis identified 14 cold spots and six low-high outliers (i.e., communes with low reported case rates surrounded by communes with high reported case rates). **Appendix 10** lists the identified areas with spatial dependence.

Figure 3.8. Local indicators of spatial association (LISA) map of average reported diphtheria case rates (per 100,000 population) in Haiti, December 2014 – June 2021



Note: The identified hotspots may, in part, reflect disparities in case detection capacity as communes with higher healthcare facility density are more likely to recognize and report cases. This potential bias should be considered when interpreting the spatial patterns.

3.4.3. Regression models

The Spearman's rank correlation analysis found that male literacy and female literacy were highly correlated ($r=0.78$, $p<0.001$). Consequently, male literacy was excluded from the pool of independent variables as it did not have a significant correlation with the reported diphtheria case rate ($p=0.18$). Low collinearity was observed among the remaining variables (VIF range=1.18–2.22).

Table 3.3 presents the results of the regression analyses. In the final OLS model, health facility density and the degree of urbanization were positively associated with the diphtheria case rate. Specifically, for every one-unit increase in healthcare facilities per 100,000 population, the diphtheria case rate was estimated to increase by 0.020. Similarly, a one-unit increase in the proportion of population who lives in urban areas led to a 0.009 increase in the diphtheria case rate. Conversely, a negative association was observed with female literacy. A one-unit increase in female literacy rate was found to decrease the diphtheria case rate by 0.030. The adjusted R^2 for the final OLS model was 0.15, which indicates that the model explains 15% of the variance in the diphtheria case rate. The R^2 value suggests a weak model fit and explanation of variance. The AIC_c score was 267.13.

Table 3.3. Summary of the ordinary least squares (OLS)[a] and geographically weighted regression (GWR)[b] models

Parameter	Initial OLS (units)	Final OLS (units)	Final GWR (units)
DTP3 coverage (%)	0.177 (0.488)		
Health facility density (per 100,000 population)	0.015 (0.007) *	0.015 (0.005) **	0.015
Improved water source (% of population)	0.003 (0.003)		
Female literacy (% of women)	-0.026 (0.007) ***	-0.024 (0.006) ***	-0.024
No toilet facility (% of households)	< -0.001 (0.004)		
School density (per 100,000 population)	-0.001 (0.002)		
Population density (people per km²)	< -0.001 (< 0.001)		
Urbanization (% of population)	0.007 (0.003) **	0.006 (0.002) **	0.006
Adjusted R^2	0.14	0.15	0.28
AIC_c	274.88	267.13	261.97

[a] For the OLS models, estimates correspond to the coefficients and the standard error in parentheses.

[b] For the GWR model, estimates correspond to the mean coefficients.

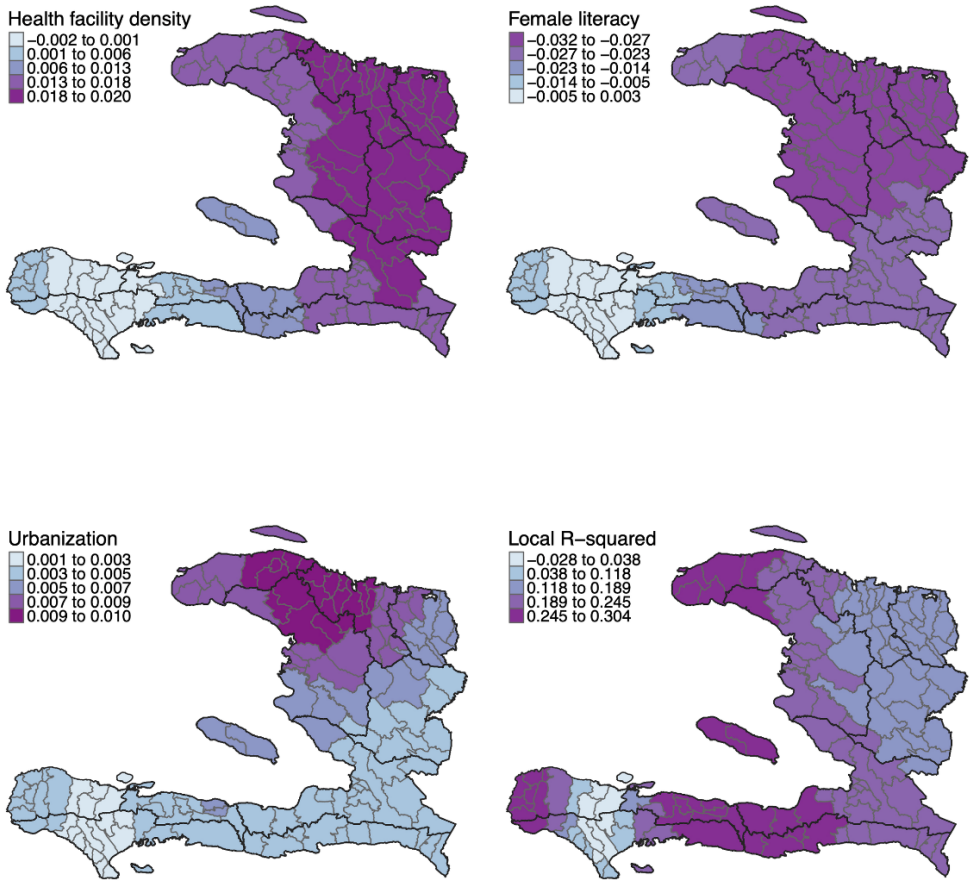
* P<0.05; ** P<0.01; *** P<0.001.

The GWR model incorporated the same variables as the final OLS model. There was agreement between the OLS and GWR model on the direction of the influence of the selected independent variables on the reported diphtheria case rate. Furthermore, the effect sizes for the independent variables were the same in the two models. However, the GWR model considerably improved model performance and fit compared to the final OLS model, as indicated by the higher adjusted

R^2 value (0.28) and lower AIC_c score (261.97). These results suggest that, by accommodating spatial non-stationarity and allowing variables to vary in space, the GWR model is better than the OLS model at explaining the relationship between the diphtheria case rate and other factors.

Figure 3.9 shows the variation in the local coefficient estimates of the GWR model and the R^2 value for each commune. These maps reveal that the influence of the three independent variables in the model varies considerably across Haiti. The local coefficients of health facility density (range=-0.002–0.020) tended to be higher in the central and northern departments of the country. The largest coefficients for female literacy (range= -0.032– -0.003) were found in Artibonite, parts of Centre and Ouest, as well as in the northern departments. Coefficients for urbanization (range=0.001–0.010) appeared to be higher in the Nord and Nord'Ouest departments and in the northernmost communes of Artibonite. The map of the local R^2 values (range=0.01–0.35) indicates that the level of explanatory power of the GWR model varies significantly throughout the territory, with higher local R^2 values found in as many as six different departments.

Figure 3.9. Local regression coefficients and R^2 values for the average annual reported diphtheria case rate (per 100,000 population) in Haiti, December 2014 – June 2021



3.5. Discussion

3.5.1. Main findings

This study has shown that the reported diphtheria case rate varied considerably between December 2014 and June 2021, reaching a peak in 2018. The investigation has identified areas with spatial dependence, which suggests that certain communes in Haiti may have predisposing factors increasing the risk of diphtheria transmission. This hypothesis is supported by findings from the GWR model, which have demonstrated that at the commune-level 28% of the variability in the diphtheria case

rate in Haiti could be explained by a combination of three factors: health facility density, the degree of urbanization, and female literacy.

The sharp increase in the diphtheria case rate in the early stages of the outbreak indicates that a large proportion of the population in Haiti was susceptible to the disease. This is consistent with the results of Minta *et al*. [301], who found no evidence of long-term protection against the infection (IgG≥1 IU/mL) among a nationally representative sample of 1,146 children aged 5–7 years in Haiti in 2017.

There are a few possible explanations for the decrease in the diphtheria case rate after 2018. That year, the MSPP conducted a mass vaccination campaign that saw more than two million children aged 1–14 years receiving at least one dose of a diphtheria vaccine [301, 302]. It is reasonable to assume that the campaign contributed to reducing the number of susceptible individuals, ultimately driving down the case rate. Nevertheless, the observed decline may have also been partly a surveillance artifact. Since 2019, there has been a dramatic surge in politically motivated protests and civil unrest, which has been accompanied by high levels of gang-related violence throughout Haiti [303]. This period has also coincided with the emergence of COVID-19 [304]. The two crises have paralyzed the country for long periods of time, making it more difficult for people in need to access medical care and for health authorities to conduct basic surveillance activities, such as case investigation and contact tracing. As a result, several diphtheria cases may have gone undetected. Therefore, available figures likely underestimate the disease's true spread.

By characterizing the spatial distribution of detected diphtheria cases, the study has confirmed the widespread transmission of the disease across Haiti. Nevertheless, substantial heterogeneities in reported case rates exist from one department to another and between communes within the same department. The LISA analysis brought to light a spectrum of diphtheria dynamics that includes several areas with spatial dependence. An estimated 646,346 people (4.7% of the population of Haiti) are living in diphtheria hotspots. Interestingly, some of the identified hotspots are located near the border with the Dominican Republic, which has reported diphtheria cases in recent years [305]. This indicates that close collaboration between the two countries, especially on cross-border surveillance, is likely to be crucial to control the transmission of diphtheria on the Hispaniola island. The hotspots detected in this study could potentially be prioritized for targeted public health interventions, including raising people's awareness about diphtheria and preventive measures through community health workers, training clinical personnel periodically, and increasing the capacity for laboratory testing. All these interventions have shown promise in the response to other public health issues in Haiti [306-308]. However, given that the full implementation of these measures will require considerable investment and time, vaccination continues to be the most vital tool in the fight against diphtheria.

It is essential, however, to interpret the findings of the LISA analysis with caution. The positive association between the reported diphtheria case rate and health facility density suggests that some communes may have been classified as hotspots, not necessarily because they truly had higher disease transmission, but rather because they possessed better access to healthcare – increasing their capacity to detect and report cases

[309, 310]. This ascertainment bias may have led to a misrepresentation of the actual geographic spread of diphtheria, potentially leaving equally or more affected areas with inadequate surveillance to appear safer than they were in reality.

Mitigating such bias requires enhancing early warning and detection mechanisms in underserved settings, alongside applying appropriate analytical corrections. For instance, in the context of tuberculosis, WHO recommends the use of inventory studies, which compare various types of data (e.g., laboratory tests, hospital admissions, pharmacy prescriptions) with official case notifications to national authorities, using record linkage to quantify under-reporting [311]. In high-burden TB settings, cases missed by routine surveillance are often detected through prevalence surveys (where a representative sample of the population is systematically tested, even if they have no symptoms or have not sought healthcare) and active case-finding (in which health workers proactively search for infected individuals in a community) [312-314]. The integration of data emerging from these approaches into the calculation of TB case rates and spatial models has helped counter the reporting imbalances introduced by unequal access to diagnostic services across different areas, allowing health authorities to distinguish between crude notification rates and more accurate estimates of actual disease frequency. Adapting these methods to diphtheria could similarly help reveal previously undetected infections, refine burden estimates towards something closer to the disease's true impact, and improve spatial models.

Beyond health facility density, the GWR model indicated that part of the variability in the reported diphtheria case rate could be attributed to the

degree of urbanization and female literacy. These associations were somewhat expected. Urban areas are generally characterized by overcrowding as well as high population mobility and inter-mixing, all of which increase the opportunities for infectious diseases, like diphtheria, to spread [315, 316]. Literate women may comprehend health messages better than illiterate women, which makes them more likely to take protective measures (e.g., vaccination and personal hygiene) for themselves and their children [227, 228]. These findings add to existing evidence that health outcomes are shaped by factors beyond healthcare [317, 318].

The coefficient estimates of the GWR model highlighted spatial variations in the relationships between the reported diphtheria case rate and the three independent variables. This suggests that the level of influence of each independent variable on the case rate may have varied from one commune to another. Gaining these local-level insights would not have been possible using global OLS techniques. These findings should be complemented by qualitative studies to understand why and how these interrelationships differ across Haiti. An example of this is the study by Chekol *et al.* [319], who detected areas with high and low contraceptive use in Ethiopia employing Getis-Ord Gi* statistic. Using GWR, they also found that higher utilization of contraceptives was associated with being married, living in urban areas, residing in affluent communities, having one to four children, having higher educational levels, and not wanting additional children. These findings were complemented by interviews, with healthcare professionals stressing the importance of accessibility and continuous education to increase awareness and acceptance of contraceptives, while community members mentioned cultural and religious beliefs, misinformation, and

long distances to healthcare facilities as barriers to the use of contraceptives. The implementation of a similar multi-pronged approach in Haiti may help to better explain the observed differences in diphtheria case rates.

Of note among the results of this study is the lack of association between the reported diphtheria case rate and risk factors related to vaccination, especially given that just 15% of the confirmed cases in this study were known to be vaccinated against diphtheria. Past research has highlighted several issues related to vaccination coverage measurements, including coverage estimates sometimes exceeding 100%, improbable year-to-year variations, and epidemics in areas reporting high coverage [320]. These issues can be linked to weaknesses in immunization information systems (IIS) and inaccuracies in vaccination coverage denominators. Unfortunately, Haiti faces both problems. A multi-country evaluation from 2009 found major flaws in the national IIS [321]. It is probable that some of those inadequacies are still present today. Furthermore, Haiti's vaccination coverage estimates are unlikely to be accurate as they are based on population projections – the last official census dates back to 2003 [322]. It is thus plausible that inadequate vaccination contributes to the propagation of diphtheria in the country, though this cannot be demonstrated through this study.

3.5.2. Limitations and strengths

Several limitations may have affected the study findings. Although diphtheria is a nationally notifiable disease in Haiti, some underreporting by physicians may still occur for a variety of reasons, including misdiagnosis. In a pooled analysis of 94 studies on diphtheria

outbreaks, Truelove *et al*. [69] found a slight negative relationship between outbreak size and CFR, suggesting that inadequate treatment of initial cases, possibly due to the incorrect identification or classification of initial patients, may have contributed to the higher mortality rates in smaller outbreaks.

Additionally, asymptomatic cases and symptomatic individuals who did not seek medical care are likely to have gone unreported. Consequently, notified cases may not necessarily reflect the true disease burden. Underreporting and misdiagnosis may have biased the analysis by generating artificially low diphtheria case rates, leading to inaccurate detection of hotspots and distorting the estimated associations between case rates and the potential risk factors. The extent of these issues may vary geographically, potentially having a greater effect on the results of communes with higher levels of underreporting and misdiagnosis.

Moreover, data for the examined variables were from different time periods. For instance, information about the number of diphtheria cases was from 2014 to 2021, data on DTP3 coverage were from 2015 to 2020, while data on female literacy, improved water source, male literacy, and no toilet facility were collected between 2016 and 2017. This mismatch in timeframes reduced the reliability of the regression estimates. Furthermore, data on certain risk factors known to correlate with diphtheria were unavailable (e.g., level of wealth, knowledge of diphtheria) [274], impeding further analysis. To mitigate these issues, if more time and resources had been available, and if Haiti's sociopolitical situation had been stable, targeted surveys and knowledge assessments could have been conducted to gather more granular data about diphtheria risk factors, further enhancing the robustness of the study.

Similarly, interviews with people affected by the disease would have allowed to obtain a more holistic picture of the epidemiology of diphtheria in Haiti.

As the models in this study were based on aggregated data, there is a risk of ecological fallacy, which is a logical error whereby associations observed at the group level are believed to necessarily hold at the individual level [323]. For instance, the detection of diphtheria hotspots may indicate that inhabitants of those communes are at an increased risk of the disease, but this does not take into consideration variation within smaller segments of the population.

Finally, like other analytic methods, GWR has some drawbacks: its spatial weighting function accounts for geographical distance but ignores the attributes of the observations [324]; local multicollinearity may be present in a GWR model, even if the independent variables are not collinear at the global level [325]. To address these shortcoming, future research on diphtheria could involve using alternative spatial analysis methods like Bayesian Hierarchical Models, which – as explained in Section 3.3.6 – can accommodate more complex relationships and better account for spatial heterogeneity.

Notwithstanding its shortcomings, this study has several strengths. The use of a subnational level dataset of diphtheria cases reported in Haiti over six years allowed for an in-depth investigation of the outbreak. By combining descriptive techniques and advanced analytical methods, a comprehensive overview of the epidemiology of the disease in the country was obtained, capturing detailed spatiotemporal patterns in reported cases, detecting persistent hotspots, and identifying potential

risk factors for the disease. The integration of data from official and open sources enhanced the robustness of the research and increased the validity of the study's conclusions.

3.5.3. Conclusions

As far as can be determined, this is the first study that describes the epidemiology of diphtheria in Haiti using GIS and spatial analysis. The study has shown that GWR is a useful technique for exploratory and descriptive data analysis, which not only improves on the OLS performance but enables the discovery of hidden spatial relationships between variables. This investigation has also demonstrated that between 2014 and 2021 diphtheria exhibited spatial variability in Haiti, with the clustering of areas with high and low reported case rates. The hotspots detected in this analysis could serve as a basis for prioritizing and targeting response activities. The baseline estimates of reported diphtheria case rates presented here could guide surveillance activities and help track progress in the control of the disease. Further research and continued monitoring of the factors found to be associated with diphtheria could help better understand the spread of the disease.

Chapter 4. The public discourse surrounding diphtheria, January 2012 – December 2022: A mixed-methods analysis using data from X (formerly Twitter)

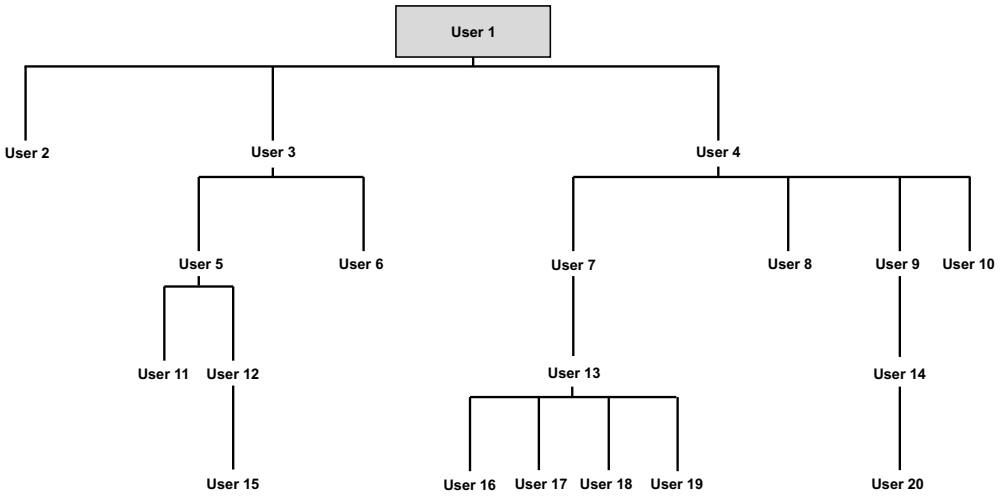
4.1. Introduction

The global persistence of diphtheria is indicative of a need to understand people's perceptions of and attitudes towards the disease and the policies implemented by governments around the world. One way of gaining this type of insight is by analyzing health-related information shared on the internet or social media, also known as infoveillance [326].

With an average of 229 million active accounts, X (formerly Twitter) is among the most prominent social media platforms worldwide [327]. X accounts often share health information, opinions, and advice through messages known as posts [328]. Information deemed important by the community propagates through reposts, a process by which posts are shared with one's followers (**Figure 4.1**). Reposting is the key mechanism for information diffusion, and this snowballing action of reposts from a follower to another means that a single message can potentially be seen by millions of people. Given the platform's reach, X data have been used to investigate acute health events, such as the 2009 H1N1 influenza pandemic [329], the 2014 Ebola virus outbreak [330], the

2015 Middle East respiratory syndrome outbreak [331], the 2015 Zika outbreak [332], and recently the COVID-19 pandemic [333].

Figure 4.1. Schematic representation of interactions on X.



- User 1 posts on X about diphtheria. His / her followers sees it.
- User 3 and User 4 see the post and choose to repost it. This expands the reach to their followers.
- Certain followers of Users 5 to 10, along some of their respective followers (Users 11 to 14), further share the post, extending its reach to additional people (Users 15 to 20).

Figure adapted from Dyar *et al*. [334]. The original figure was modified for clarity and additional context.

It was assumed that examining diphtheria-related posts might help to understand how the public perceives the disease and the interventions implemented by governments. Such analysis might also further elucidate social media's role in shaping people's opinion and attitudes towards vaccine-preventable diseases. Therefore, the aim of this research was to investigate the discourse surrounding diphtheria on X. Specifically, the study sought to determine the volume and characteristics of posts relating to diphtheria, explore spatiotemporal trends of diphtheria discussions, identify the main topics emerging from the conversations, and discover the key themes and patterns in the data.

4.2. Methods

4.2.1. Study design

This study involved a mixture of quantitative and qualitative methods to analyze information contained in X posts relating to diphtheria. Quantitative methods used in this study included descriptive and statistical methods to perform spatiotemporal, term frequency, and hierarchical clustering analyses [335]. The qualitative component of the study was based on Grounded Theory, which is a research method that systematically integrates data collection with concurrent analysis to iteratively construct theories directly from the empirical evidence, rather than from preconceived ideas [124, 336]. While the quantitative methods sought to discern the measurable patterns in diphtheria-related posts, the qualitative analysis was used to understand the content of these posts. Ultimately, this mixed-methods approach provided a nuanced understanding of the public discourse on diphtheria.

4.2.2. Data collection

Data for this study was collected and compiled by researchers at Boston Children's Hospital (BCH), affiliated with Harvard University, possessing access to historical posts published on X via a pre-existing agreement with the social media platform. The data were accessed on 15 July 2023. X's application programming interface (API) was used to identify posts containing the term "diphtheria" that had been published between 1st January 2012 and 31st December 2022. This period was chosen as it had been characterized by a global resurgence of diphtheria alongside a notable growth in digital communication, especially on social media platforms like X, offering a timely context to study trends in the diphtheria-related discourse.

In line with the API limitations, collected posts represented a random sample (less than 1%) of all public posts published on X containing the term "diphtheria" during the study period. The data were provided in multiple CSV files, with each line representing a distinct post. Collected variables included post ID (i.e., a unique numerical string in the URL of any post), post text, date of publication, geographical coordinates, usernames, bio (i.e., a brief description of an account's identity or brand on X), number of followers, and number of accounts followed (**Appendix 11**).

4.2.3. Data preprocessing

All data preprocessing and quantitative analyses were conducted using R language [337] through the RStudio version 2022.07.1 interface [338]. The study's dataset included replies as well as posts from unverified accounts. Furthermore, no software application was used to identify posts published by bots (i.e., automated accounts that can imitate or replace human behaviour) as they also contributed to the conversations surrounding diphtheria.

To prepare the collected data for the quantitative analyses, all reposts, duplicate posts, and unintelligible posts (e.g., those containing only the word "diphtheria" and no additional terms) were manually reviewed and removed. To align with the linguistic expertise available within the research team and ensure consistency in the analysis, posts in languages other than English were also filtered out. Unfortunately, the exact number of non-English posts excluded was not recorded. Subsequently, all letters in the posts were converted to lowercase to ensure that the same words with different cases (e.g., "Vaccine" and "vaccine") were

treated analogously by the algorithms used in the quantitative analyses. Similarly, stemming – a process in which each word is changed to its root form (e.g., "information" and "informative" are transformed to "inform") – was performed so that different variations of the same words were analyzed together by the algorithms [339]. Then, emoticons, extra white space, punctuation, symbols, URLs, single-letter words, and stop words (i.e., commonly used words of little analytical value, such as "the", "it", "for") were removed to minimize noise in the analyses [339]. The word "diphtheria" was filtered out to ease the identification of key terms and topics. Finally, the text of posts was split into smaller parts called tokens, after which a Document-Term-Matrix (DTM) (a table in which rows correspond to posts in the dataset, while columns represent terms in the posts) was built to facilitate the analysis by the algorithms [339].

For the qualitative analysis, instead of conducting extensive data preprocessing, the original content of the posts that were retained following the removal of reposts, duplicate posts, and unintelligible posts was examined. By analyzing the unaltered posts, the intention was to preserve the raw voices and authentic expressions specific to the public discourse surrounding diphtheria. This minimally invasive approach was intended to capture as accurately as possible the genuine opinions and sentiments of X users, providing a rich ground for a new theory to emerge while enhancing the validity of the analysis.

4.2.4. Descriptive analysis

To understand the characteristics of X accounts, the mean, median, and interquartile ranges were estimated for key variables, such as the

number of followers and followed accounts. Furthermore, the 100 accounts with the highest number of followers were grouped into different categories, based on keywords and phrases provided in their X bios. Accounts that could fit into multiple categories were categorized based on the most dominant theme in the bios. Accounts with no bios were not categorized.

4.2.5. Spatiotemporal analysis

Using timestamps and geolocation data, the distribution of diphtheria-related posts was analyzed across space and time. Firstly, a dot map was developed to visualize the geographic spread of the posts. To better understand spatial trends, countries were categorized by WHO region [340] and World Bank's income levels [341].

Subsequently, to investigate temporal trends in diphtheria conversations, a plot was generated, showing diphtheria-related posts and cases reported annually to WHO between 2012 and 2022 [59].

4.2.6. Statistical analysis

The Spearman's correlation test [294] was performed to examine whether there was an association between diphtheria-related posts and globally-reported cases. The analysis provided a coefficient that ranged from -1 to 1, with a positive value indicating a direct correlation, a negative value denoting an inverse correlation, and a value close to zero implying little or no correlation.

A cross-correlation analysis [342, 343] was conducted to determine whether the posts and cases were correlated over time. The analysis

generated a set of lag values by systematically shifting one dataset forward or backward in time relative to the other. The lag values were accompanied by the autocorrelation function (ACF) values, which were the corresponding coefficients indicating how the two datasets interacted at defined time lags. Similar to the Spearman's correlation test, the ACF ranged from -1 to 1, with positive values representing direct correlations, negative values suggesting inverse correlations, and values near zero revealing little to no correlation. A cross-correlation function (CCF) plot was produced, illustrating these correlations at different lags. In the plot, the ACF values were presented as vertical bars, with those above the x axis indicating direct correlations and those below implying inverse correlations. The length of the bars increased with the strength of the correlations, with bars exceeding confidence intervals (represented on the plot as horizontal dotted lines above and below the zero line) denoting statistically significant correlations. The level of significance was set at $p<0.05$.

4.2.7. Term frequency analysis

To obtain an initial overview of the recurrent topics associated with diphtheria, the most frequently used terms in the posts were identified by counting the occurrence of each word in the dataset. Estimated frequencies were visualized in a bar chart and a word cloud.

4.2.8. Hierarchical clustering

To enhance the understanding of the key topics of discussion, hierarchical clustering was used, which is an algorithm that groups data points in a collection of documents into semantic word clusters (called "topics") based on their similarity [335]. Specifically, the agglomerative clustering algorithm with the Ward linkage method [344] was applied

because it minimizes the total within-cluster variance – resulting in clusters that are more compact and defined compared to other methods, such as single linkage (in which clusters are combined based on the shortest distance between any two elements in the clusters), complete linkage (in which clusters are merged according to the longest distance between any two elements in the clusters), and average linkage (in which clusters are joined using the average distance between all pairs of elements in the clusters). The analysis produced a dendrogram (i.e., a tree-like structure that highlights emerging topics by showing the clustering of key words). In the dendrogram, the closeness of terms indicated their proximity in the dataset.

4.2.9. Grounded theory analysis

Originally developed in the 1960s by sociologists Barney Glaser and Anselm Strauss, Grounded Theory is a systematic, yet flexible, method for inductively building theories from qualitative data [345]. It is used to understand a phenomenon through the lens of those who experience it. Contrary to qualitative methods that typically rely on established categories or frameworks, such as content analysis and discourse analysis, Grounded Theory generates hypotheses and theories after data collection to ensure that findings are not influenced by preconceived notions [346, 347]. It is particularly useful for the investigation of complex phenomena, for which there is little or no existing theory [347, 348]. Considering the limited literature on diphtheria, Grounded Theory appeared to be a fitting approach for this inquiry.

To obtain a global understanding of people's perceptions of and attitudes towards diphtheria, posts from countries with high and low

burden of diphtheria were sampled and analyzed separately. The process began by purposively sampling the original posts published from countries with a high burden of the disease. The assumption was that information from these areas could provide crucial insights into the knowledge, beliefs, and attitudes of those most affected by the disease. In this analysis, a "high-burden country" was defined as having reported over 200 diphtheria cases between 2012 and 2022 based on WHO data [59]. This threshold represents a significant share (37%) of the average cumulative number of cases reported per country during the study period (536), helping to differentiate between countries with notable levels of transmission and those with sporadic outbreaks. Furthermore, it accounted for variations in reporting accuracy and completeness, ensuring that countries with a significant but underreported diphtheria burden were not excluded. Despite its arbitrary nature, the chosen threshold enabled the compilation of a manageable yet comprehensive sample of posts from countries where diphtheria had represented a sustained public health challenge.

Posts from "low-burden countries" were sampled using the day of posting as the primary matching criterion. If a post from a "high-burden country" corresponded to multiple posts from "low-burden countries", the post from the account with the largest number of followers was selected. This approach ensured temporal alignment across the two samples and inclusion of posts with the greatest potential reach.

In accordance with Grounded Theory principles, the posts were systematically analyzed using open coding, whereby posts were closely examined to identify emerging codes [123, 336]. During this process, three members of the research team independently assigned codes to

each post to facilitate the detection of patterns or variations in the data. Discrepancies in code assignment were reconciled through team discussions. If needed, the URL of a post was accessed to better understand the context.

Subsequently, axial coding [123, 336] was performed, during which codes relating to similar concepts were grouped into categories, leading to the formation of potential subthemes. Lastly, selective coding was conducted, in which identified subthemes were related to one another and to the main themes in the dataset to form a cohesive theory [349].

Once finalized, the theoretical framework was visually presented using a mind map that outlined the linkages among identified themes and subthemes. To facilitate the analysis, Microsoft Excel (Microsoft, Redmond, United States) [350] was used.

4.2.10. Researcher positionality

Recognition of researcher positionality is critical to qualitative inquiry, as it highlights the subtle ways in which the investigator's own identity, social environment, and personal history can profoundly impact the design, analysis, and findings of a study [351-353]. The nearly three years I spent in Haiti as an epidemiologist for PAHO / WHO provided me with some firsthand insights into the challenges represented by diphtheria in such contexts. Nevertheless, I – like most members of the research team – originated from a country where the disease is rarely encountered. As a result, we may not have fully appreciated the everyday realities of individuals in high-burden settings, potentially affecting the way we interpreted their X posts [351].

The research team consisted of experts in economics, epidemiology, statistics, and veterinary medicine, who all attended Western universities. Our academic training, which focused predominantly on quantitative principles and techniques, possibly guided both our selection of analytical approaches and the relative value we placed on different sets of information [354]. Additionally, the team included three female supervisors, one male supervisor, and one male PhD candidate. Gender dynamics influenced by societal norms and expectations may have subtly shaped how we interpreted posts, how we resolved differences of opinion, and how we framed our conclusions [129, 355]. Furthermore, our prior assumptions and stance on vaccination and public health likely informed how we identified themes, particularly those involving vaccine hesitancy and misinformation [356]. Finally, because of our language expertise, we confined our study to only posts in English, inevitably omitting perspectives from populations that communicate in other languages [357].

To counteract these biases, we systematically practiced reflexivity, examining our assumptions and methods throughout the research process [352]. By gathering posts from countries with both high and low diphtheria burden, we attempted to reflect differing realities and capture multiple viewpoints. These posts were reviewed and coded collaboratively, with team members openly challenging one another's assumptions and interpretations through open dialogue. Furthermore, findings from the Grounded Theory analysis were triangulated with the results of various descriptive and statistical methods to increase the validity of the research [356, 358]. Despite these efforts, we acknowledge that our understanding of the public discourse surrounding diphtheria still remain somewhat shaped by our positionality.

4.2.11. Ethical consideration

The study was approved by the University of Nottingham's School of Medicine Research Ethics Committee (reference number: FMHS 38-0722) (**Appendix 12**). X's terms of service included the possibility of posts being used for research purposes. Data collection was conducted through X's API, which prevents the extraction of any private and protected information. Thus, only data publicly shared by X accounts were collected for this study. Since the study did not involve direct engagement with human participants, individual consent was not collected. Given that the data were publicly available and anonymized, informed consent was not required as per the Research Ethics Committee's decision. All personally identifiable information (e.g., post IDs and usernames) was stored securely and excluded from any publication related to this study. To further protect user privacy, direct quotes from the posts were omitted.

4.3. Results

4.3.1. Descriptive analysis

A total of 2,916 diphtheria-related posts published between 1st January 2012 and 31st December 2022 were extracted. After the removal of reposts, duplicate posts, and unintelligible posts, 2,452 posts were retained.

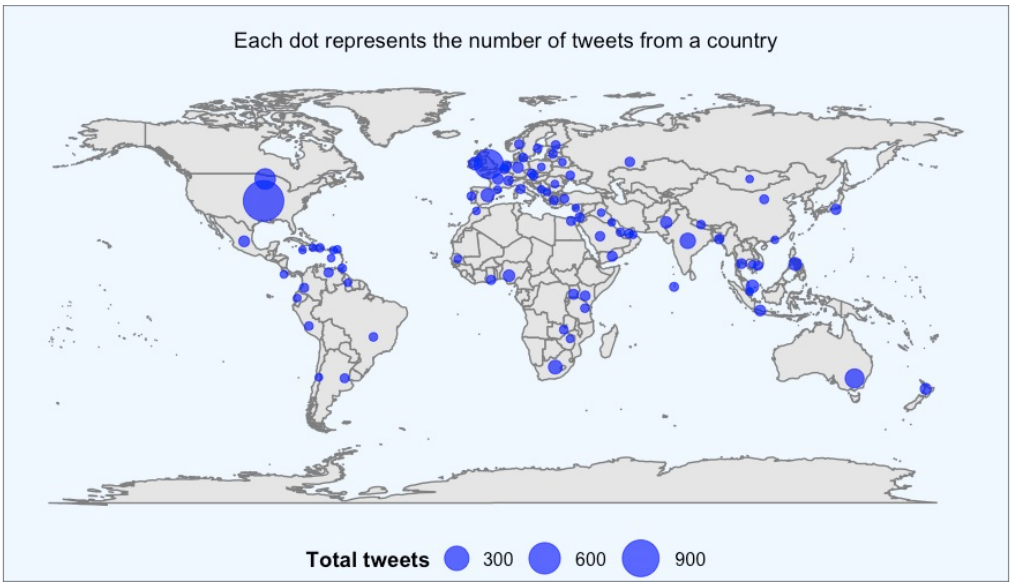
During the study period, 1,022 unique accounts tweeted about diphtheria. Their follower count ranged from 0 to 4,863,994. The mean and median number of followers was 6,624 and 710, respectively, with an interquartile range of 200 to 1,882. Notably, the follower count was missing for 21 posts.

The top 100 most-followed accounts in the study's sample were classified into five groups. The most prominent group at 27% consisted of Political Actors, including elected officials and social advocates. About 23% were Lifestyle Enthusiasts, who were interested in topics such as fitness, hobbies, and parenting. Approximately 21% were Media Corporations and Personalities, including news outlets, journalists, and content creators. Medical Institutions and Experts constituted 15% of the accounts. Lastly, Other Individuals and Groups who did not fit into the above categories represented 14% of the accounts.

4.3.2. Spatiotemporal analysis

Spatial data were available for most posts (2,440; 99.5%), indicating that these had been published from 84 countries across all six WHO regions (**Figure 4.2; Appendix 13**). Despite the widespread geographical distribution, most posts originated from the Americas (1,363; 55.9%) and Europe (629; 25.8%). The rest came from the Western Pacific (226; 9.3%), South-East Asia (95; 3.9%), Africa (82; 3.4%), and Eastern Mediterranean (45; 1.8%) regions. The Americas and Europe represented 3.1% (3,116 / 101,928) of diphtheria cases reported during the study period [59].

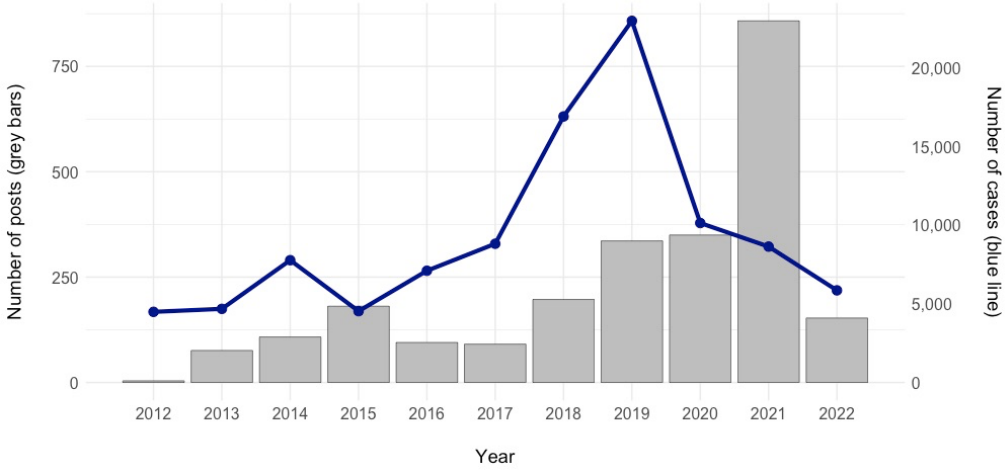
Figure 4.2. Geographic distribution of X posts mentioning the term "diphtheria", 2012 – 2022.



The distribution of posts also revealed a concentration in high-income countries (2,130; 87.3%). This was followed by lower-middle (141; 5.8%), upper-middle (128; 5.3%), and low-income countries (41; 1.7%). Among low-income countries, only one post came from Haiti. Throughout the study period, mobile cellular subscriptions in Haiti remained relatively low and only rose marginally from 60 to 64 per 100 people between 2012 and 2021 – well below the global average of over 100 subscriptions per 100 people [359]. Importantly, high-income countries accounted for just 0.9% (896 / 101,928) of diphtheria cases notified during the study period [59].

The volume of diphtheria-related posts rose over time, going from four to 858 posts between 2012 and 2021 (**Figure 4.3**). By 2022, the number of posts declined to 153. During the same period, the global number of diphtheria cases increased from 4,490 in 2014 to 22,986 in 2019, decreasing to 5,856 by 2022.

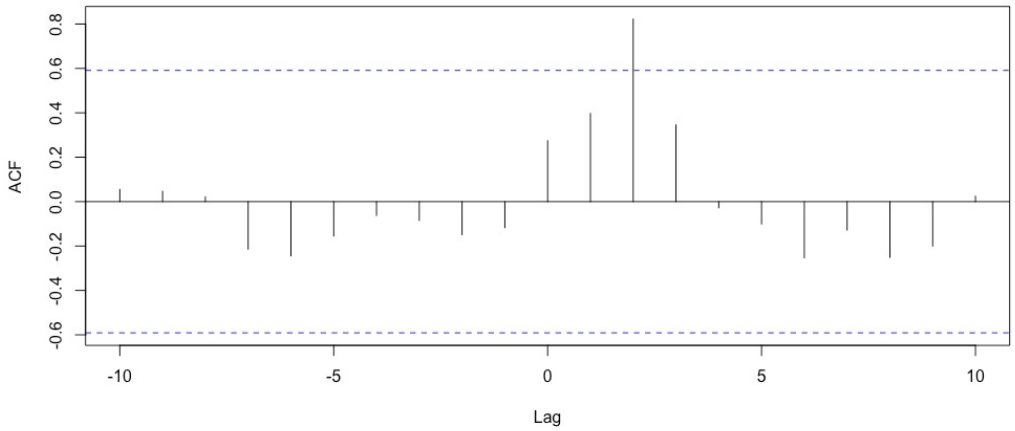
Figure 4.3. Number of diphtheria cases and X posts mentioning the term "diphtheria", 2012 – 2022.



The calculated Spearman's correlation coefficient of 0.627 ($p<0.05$) denoted a strong, statistically significant positive association between the number of diphtheria-related posts and the number of diphtheria cases reported globally.

As illustrated by the CCF plot, the cross-correlation peaked at lag +2, suggesting that posts led diphtheria cases by two years (**Figure 4.4**). The other cross-correlations were not statistically significant as shown by the bars appearing within the confidence intervals.

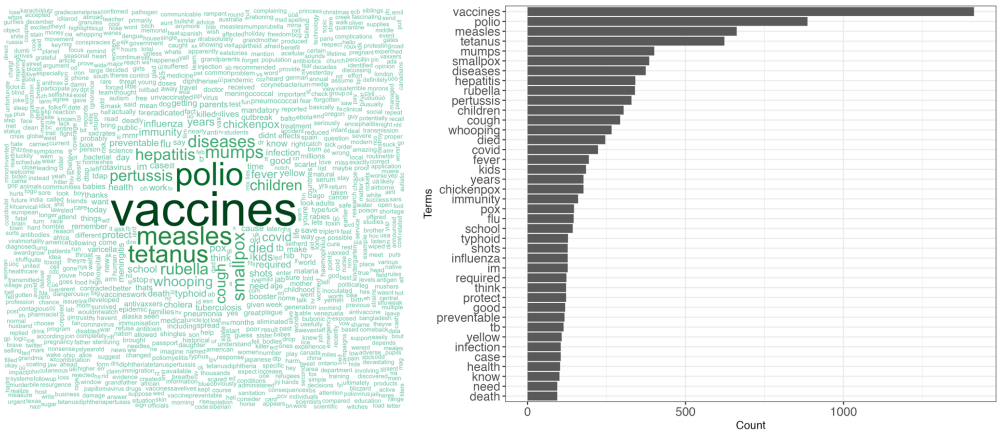
Figure 4.4. Cross-correlation of yearly diphtheria-related X posts and diphtheria cases, 2012 – 2022.



4.3.2. Term frequency analysis

After excluding the word "diphtheria", URLs, single-letter words, stop words, numbers, symbols, and emoticons, the dataset had 5,291 unique terms. The most common term in the dataset was "vaccines", with a count of 1,413 (**Figure 4.5**).

Figure 4.5. Word cloud and bar chart of the most common words in diphtheria-related X posts, 2012 – 2022.

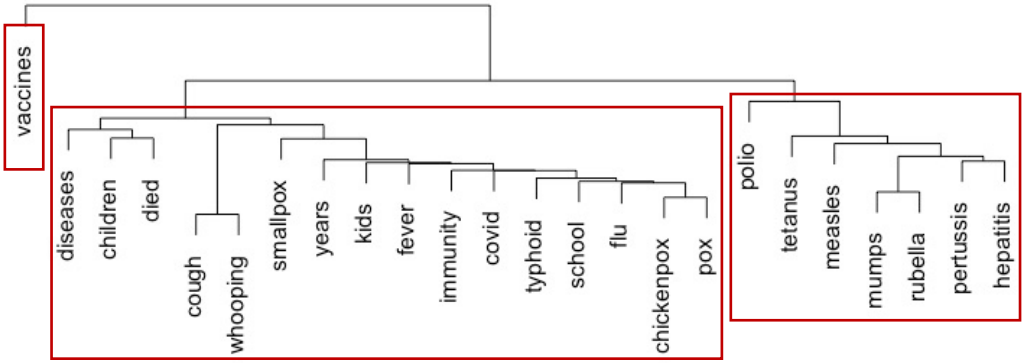


Other frequently mentioned terms included “polio” (886), “measles” (662), “tetanus” (623), “mumps” (401), “smallpox” (385), “hepatitis” (341), rubella (340), pertussis (329), and “covid” (223). Additional terms with high recurrence were “children” (304), “kids” (185), and “school” (143). The dataset also featured terms such as “died” (247), “people” (197), and “fever” (194).

4.3.3. Hierarchical clustering

The hierarchical clustering analysis produced a dendrogram, comprising three main clusters, each with different terms (**Figure 4.6**). The first cluster was called “Vaccination” because the sole term in this cluster was “vaccines”. The second cluster was named “Health impact” as it included terms relating to medical illnesses (i.e., “diseases”, “whooping”, “cough”, “smallpox”, “covid”, “typhoid”, “chickenpox”, “pox”), consequences of disease spread (i.e., “died”, “fever”, “immunity”), and affected demographics (i.e., “children”, “kids”, “schools”). The third cluster was labeled “Vaccine-preventable diseases” given that all seven terms in this cluster were diseases that could be prevented through vaccination (i.e., “polio”, “tetanus”, “measles”, “mumps”, “rubella”, “hepatitis”, “pertussis”).

Figure 4.6. Cluster dendrogram of topics emerging from diphtheria-related X posts, 2012 – 2022.

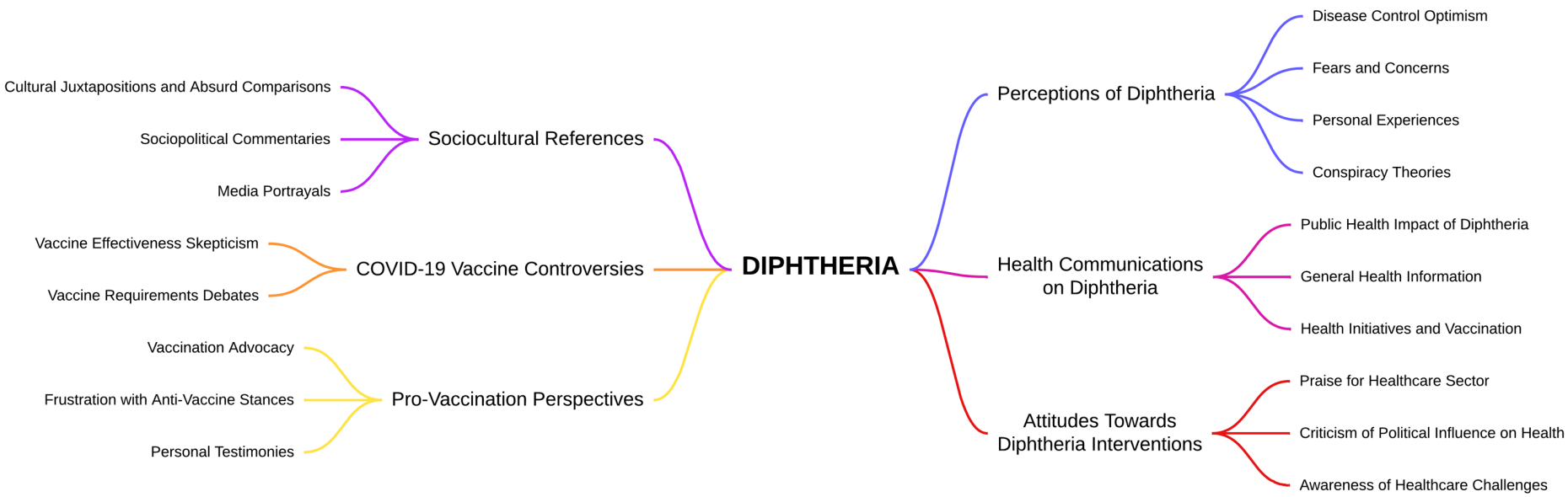


4.3.4. Grounded theory analysis

The sample for the Grounded Theory analysis included 163 posts from 10 high-burden countries and 163 posts from 25 low-burden countries (**Appendix 14**). All six WHO regions were represented.

The examination of these posts identified six themes: Perceptions of Diphtheria, Health Communications on Diphtheria, Attitudes Towards Diphtheria Interventions, Pro-Vaccination Perspectives, COVID-19 Vaccine Controversies, and Sociocultural References. These were further divided into 18 subthemes (**Figure 4.7**). To safeguard user privacy, quotes from the posts were not included.

Figure 4.7. Themes and subthemes emerging from diphtheria-related X posts, 2012 – 2022.



Perceptions of Diphtheria

The Perceptions of Diphtheria theme highlighted a dichotomy of hope and fear towards the disease. Posts discussing declines in reported diphtheria cases, including its near eradication in both low-burden countries like the United States (US) and high-burden countries such as Nepal, revealed a positive expectation about the containment of the illness (Disease Control Optimism). These public health achievements were credited to vaccines and medical breakthroughs, with one post citing Alexander Fleming for his contribution to the development of penicillin.

Conversely, a greater number of posts expressed concern about the reemergence of vaccine-preventable diseases like diphtheria – with X users in India and the Philippines ascribing it to the countries' low vaccination rates, while users in Yemen attributed it to the ongoing conflict (Fears and Concerns). Other posts showed apprehension about the potential resurgence of the disease in Europe and North America, with some linking it to immigration.

The severity of diphtheria was further underlined by first-hand accounts of ordinary citizens in predominantly high-burden countries, who discussed the deaths of family members, while healthcare professionals reflected on the difficulties of managing patients, often resulting in the loss of young children (Personal Experiences).

Finally, a small segment of the discourse focused on unfounded speculations – with a post claiming that diphtheria deaths were orchestrated by a "vaccine mafia", while others erroneously associated

the disease with the passing of historical figures (e.g., Hitler's mother, Princess Alice of Battenberg) (Conspiracy Theories).

Health Communications on Diphtheria

Posts under the Health Communication theme disseminated information and asked questions about a wide range of topics relating to public health and diphtheria. Several posts alerted about the detection of imported cases in low-burden settings and ongoing outbreaks in countries like India, the Philippines, and Yemen – often sharing real-time updates on reported cases, hospitalizations, and deaths (Public Health Impact of Diphtheria). The tone was primarily neutral, becoming emotionally charged when focusing on certain topics, such as how diphtheria is "still killing kids". Notably, several posts linked the return of diphtheria to declines in childhood immunization during the COVID-19 pandemic.

The posts also mentioned diphtheria together with other diseases, discussing their causes, symptoms, treatments, and preventive strategies (General Health Information). Some of these posts were published by healthcare professionals, who used technical language to share information about pathogens or academic literature.

Additionally, certain posts provided information about ongoing vaccination campaigns against diphtheria and other diseases in countries such as Haiti, Indonesia, and Thailand, as well as other health interventions, including a community survey in India, vaccine development efforts in Canada and Cuba, a maternity healthcare plan in South Africa, and an online app to prepare for medical exams (Health Initiatives and Vaccination).

Attitudes Towards Diphtheria Interventions

The Attitudes Towards Diphtheria Interventions theme provided insights into people's opinions of health measures implemented against diphtheria. Posts from Bangladesh, India, Pakistan, Ukraine, and the United Kingdom lauded public health programmes and vaccination initiatives for their contributions in disease prevention and control (Praise for Healthcare Sector).

Other posts highlighted negative perceptions of the impact of politics on public health (Criticism of Political Influence on Health). For instance, some posts attributed the spread of diseases in Yemen to military actions perpetrated by the US and Saudi Arabia. Others challenged the narrative that the political leadership in India, specifically Prime Minister Narendra Modi and his Bharatiya Janata Party, should receive significant credit for the production and distribution of vaccines in the country, emphasizing instead the critical roles played by established scientific, industrial, and medical institutions.

Attention was also drawn to problems encountered in India, Pakistan, and Yemen during the response to diphtheria outbreaks, including inadequate patient referral systems, lack of isolation units, alongside stockouts of vaccines and diphtheria antitoxin (Awareness of Healthcare Challenges). Crucially, several of these posts urgently called for medical assistance.

Pro-Vaccination Perspectives

The Pro-Vaccination Perspectives theme mainly promoted the importance of vaccination, stressing its vital role for the prevention and control of diphtheria and several other diseases (Vaccination Advocacy).

These posts often presented historical lessons or statistics on children saved through immunization. Some posts called for vaccines to be made mandatory, while others praised World Immunization Day – with one post from Pakistan describing vaccines as "one of the great triumphs of modern medicine".

In both high- and low burden countries, the pro-vaccination sentiment was frequently underscored by resentment towards anti-vaxxers – as demonstrated by posts criticizing their ideologies, rebutting their arguments, and highlighting the disease control challenges posed by their attitudes and (in some cases) religious beliefs (Frustration with Anti-Vaccine Stances). Anti-vaccination views were regularly challenged using sarcasm and rhetorical questions, especially in low-burden countries.

People were repeatedly urged to get vaccinated through personal anecdotes of healthcare workers and parents, which highlighted a plethora of emotions – including the fear of needles, hope for no side effects, and solace for the immunity conferred by vaccination (Personal Testimonies).

COVID-19 Vaccine Controversies

The COVID-19 Vaccine Controversies theme revealed a distrust of COVID-19 vaccines. All posts under this theme originated from low-burden countries, namely Australia, Canada, and the US. Several posts raised concerns regarding the perceived low effectiveness of COVID-19 vaccines to prevent disease, with comparisons being made to the ability of traditional vaccines to eliminate diseases like diphtheria and polio (Vaccine Effectiveness Skepticism).

Conversations also touched upon vaccine mandates for COVID-19 (Vaccine Requirements Debates). While one post questioned the legitimacy of vaccine passports for COVID-19 given the disease's "survival rate of 99.9%", two other posts presented diametrically opposing views regarding a proposed law that would bar schools in the US state of Ohio from requiring vaccinations.

Sociocultural References

Posts under the Sociocultural References theme often combined cultural narratives, humour, and satire to comment on current issues. Several posts, which were predominantly from low-burden countries, trivialized diphtheria – using it as part of jokes, comparing it hyperbolically to other diseases, or linking it to regional stereotypes (Cultural Juxtapositions and Absurd Comparisons).

Diphtheria was also cited metaphorically to voice discontent with how critical sociopolitical issues are being discussed and managed (Sociopolitical Commentaries). One post sardonically commented on the economic burden of diphtheria treatment, likening the situation to a "communist society" where financial responsibilities are unfairly distributed, with only a few people bearing the bulk of the costs. Another post satirized the perceived excessive blame placed on the American politician Hillary Clinton, who was held responsible for various societal problems, including those as unrelated as diphtheria. A third post decried the decline of logical reasoning in public dialogue, drawing on exaggerated imagery such as diphtheria and typhus from California street feces to depict the decay of rationality.

Finally, some posts presented passionate reactions to the portrayal of diphtheria in the movie "Togo" and the TV show "Call the Midwife", while another post mentioned the disease as the subject of an upcoming podcast (Media Portrayals).

4.4. Discussion

4.4.1. Main findings

Through a comprehensive methodological framework, this study collected and examined a large dataset of X posts, employing various quantitative and qualitative techniques. The adopted approach revealed that a diversity of voices participates in conversations around diphtheria on X, leading to the emergence of a variety of themes and subthemes – with vaccination being the predominant topic of interest. The analysis also highlighted differences in tone and content among countries, mirroring variations in healthcare priorities and challenges.

This study gathered almost 3,000 diphtheria-related posts published on X between 2012 and 2022. While this figure is significant, past studies have shown considerably higher levels of engagement on X for other diseases. For instance, Meadows *et al*. [360] found more than 1 million posts containing the word "measles" from December 2014 to April 2015. Househ [361] discovered that there were approximately 26 million posts that included the word "ebola" in October 2014 alone. The lower number of diphtheria-related posts suggests that the disease may not be as prominent in social media discourse as other illnesses. This could be due to diphtheria being perceived globally as a lesser threat to human health compared to other diseases with higher infectivity, pathogenicity, or virulence. It is also important to note that the observed differences in

the number of posts may be partly explained by other studies having employed more comprehensive methods for the collection of posts or having had access to more extensive datasets.

During the study period, over 1,000 unique accounts posted about diphtheria. Their follower counts varied widely, with a minority of accounts possessing a disproportionately large number of followers. Similar skewed distributions have been observed in other analyses. In a study of 41.7 million X accounts, Kwak *et al*. [362] found that only 40 accounts (mainly public figures and media companies) had over a million followers. Likewise, using data from nearly 6,000 X accounts, Zhu and Lerman [363] concluded that the top 20% of accounts (based on follower counts) comprised more than 96% of all followers.

This study revealed different contributors to diphtheria-related discussions on X, including Political Actors, Lifestyle Enthusiasts, Media Corporations and Personalities, Medical Institutions and Experts, and Other Individuals and Groups. This finding is consistent with previous studies [364, 365], which also found a variety of accounts involved in online conversations relating to health issues. The variety of perspectives observed in the present study suggests that public awareness about diphtheria transcended the healthcare sector, extending to other societal domains. This may be due to the extensive risk and impact of the disease.

It is important to highlight that only a small portion of posts was shared by Medical Institutions and Experts. Porat *et al*. [366] also found that none of the popular posts published on X following the detection of a diphtheria case in Spain in 2015 came from health organizations. These

findings are in contrast with those of several other studies [367-369], which revealed how health organizations like WHO were central to X conversations on COVID-19, particularly during the early stages of the pandemic, after which the influence of health organizations began to wane with the rising prominence of political figures and celebrities. The increased engagement of health organizations during the COVID-19 pandemic compared to recent diphtheria outbreaks may be due to the unprecedented impact and scale of the COVID-19 emergency, which compelled health organizations to continuously share authoritative, accurate, and up-to-date information about the disease with the global community.

X posts included in this study originated from 84 countries across all WHO regions, suggesting that there may be global interest in diphtheria. However, the geographical distribution of these posts did not align with the global burden of the disease. For instance, Haiti accounted for just a single post, despite reporting a high number of cases. Notably, most posts originated from high-income countries in the Americas and Europe, even though these settings had the lowest case counts. This result differs from those of past studies on Zika and Lyme disease [332, 370], which reported correlations between the geographical distribution of cases and X posts. The predominance of high-income countries in diphtheria-related conversations on X could partly be reflective of their increased web access and social media use, considering for instance that 93% of the world's population not connected to mobile internet resides in LMICs [53]. It is also possible that expanding the analysis to other social media platforms that are perhaps more popular in LMICs, such as Facebook [371], would have resulted in a higher volume of diphtheria-related posts from these settings.

This analysis highlighted an increasing trend in diphtheria-related discussions over time, with a spike in 2021. This peak coincided with the global rollout of COVID-19 vaccines, which may have increased public interest on vaccination and vaccine-preventable diseases. This broadly supports evidence from previous studies that found changes in social media activity after major vaccine-related announcements during the COVID-19 pandemic [372, 373]. Meanwhile, the surge in the volume of posts could have been partly due to heightened public awareness or concern about diphtheria because of its global resurgence. Importantly, the rising trend in the volume of posts could have also been due to the growing popularity of X and availability of smartphones over the years [374].

In this study, the Spearman's correlation and cross-correlation analyses pointed to a potential positive association between diphtheria-related posts and cases, with increases in global posts preceding actual surges in cases reported worldwide by two years. These findings should be interpreted with caution as the two-year lag observed in this study could have been coincidental or determined by numerous other factors that influence posting behaviour, such as changes in vaccination policies, societal attitudes towards diseases and vaccines, and media coverage of health emergencies. The need for cautiousness becomes even more evident when findings from this research are compared with those of other studies relating to diseases like COVID-19 and avian influenza [375-378], which reported significantly shorter lags between increases in X posts and surges in cases ranging from 1 to 15 days.

Despite the longer time lag compared to other studies, findings from the cross-correlation analysis are still useful. For instance, they enabled the

identification of a possible statistical relationship between seemingly unrelated variables like X posts and reported diphtheria cases that warrants further investigation. This potential link between discussions on X and disease trends may simply have been missed using traditional surveillance methods. If corroborated by future studies, surges in diphtheria-related posts could serve as early indicators of emerging outbreaks. Further research should utilize data at a finer timescale, which could facilitate the examination of trends at a daily, weekly, or monthly level, ultimately helping to refine the observed time lag. It may also be interesting to explore whether the time lag varies geographically, potentially reflecting differences across countries in social media usage and reported diphtheria rates.

The term frequency and hierarchical clustering analyses conducted as part of this study indicated that conversations on X surrounding diphtheria frequently went beyond the disease itself, addressing other vaccine-preventable diseases and focusing on illness prevention through vaccination. Other studies investigating social media discussions on infectious diseases also found a variety of topics emerging from the posts. For instance, Bogdanowicz and Guan [379] collected X posts in April 2020 to understand the evolving discourse around COVID-19 in the United States, identifying 12 prevalent topics that encompassed politics, healthcare, community, and the economy. Similarly, de Melo and Figueiredo [380] analyzed X posts about COVID-19 in Brazil published from January to May 2020, discovering 20 key topics that covered personal stories, political issues, economic effects, medical treatment and research, prevention and control, and the impact of the pandemic on entertainment. Collectively, these findings reveal a

common pattern where conversations on a specific disease spark discussions relating to wider societal concerns.

While the term frequency and hierarchical clustering analyses provided a surface-level overview of the topics discussed on X in relation to diphtheria, Grounded Theory allowed to obtain deeper understandings of the themes, subthemes, and emotions emerging from the posts. The analysis showed that several posts did not specifically focus on diphtheria but rather mentioned it alongside other vaccine-preventable diseases. The dominant narrative stressed the importance of immunization for the prevention and control of infectious diseases, with several accounts providing reasons to get vaccinated and expressing gratitude towards the healthcare sector. Personal testimonies offered unique insights into the impact of diphtheria, reinforcing the significance of preventive measures. These pro-vaccination perspectives were challenged only by a minority of posts, including some that spread misinformation and conspiracy theories. The low proportion of posts containing this type of content may partly be explained by X's reinforcement of policies against misinformation in response to the COVID-19 pandemic, resulting in the removal of harmful and misleading posts starting from 2020 [381]. This may have created a temporal bias in this study's dataset, potentially hindering further insights into the public discourse surrounding diphtheria and the extent of misinformation. Crucially, a segment of the conversation focused on the introduction of COVID-19 vaccines, with the prevailing sentiment being negative and several posts expressing doubts regarding their effectiveness. Other concerns that emerged from previous research on X include the fear of side effects following COVID-19 vaccination [382-384].

Overall, the Grounded Theory analysis painted a picture of a community that proactively pursued and disseminated knowledge on diphtheria, from its pathophysiology to preventive strategies, highlighting the role of X in educating and informing the public. Differences in tone and content between countries were also observed. Posts from high-burden countries repeatedly expressed frustration with systemic healthcare challenges and inadequate outbreak responses, demonstrating a direct and urgent engagement with a disease that was seen as an immediate threat. In contrast, posts from low-burden countries adopted a more reflective stance, highlighting historical lessons, deliberating on vaccine policies, and addressing more broadly disease prevention. Moreover, posts from low-burden countries often included humour and sarcasm, suggesting a more distant and abstract relationship with diphtheria.

As illustrated above, through the application of Grounded Theory, it became possible to gain comprehensive insights into the public discussions about diphtheria on X, which may not have been evident through other qualitative methods. For example, Pulido *et al*. [385] employed content analysis to investigate false and evidence-based information on X during the early days of the COVID-19 pandemic. By combining predefined and emergent categories, the authors found that posts containing false information about COVID-19 (e.g., rumors, conspiracy theories, and misleading videos of people collapsing) were published more than science-based evidence posts (e.g., proper mask usage and SARS-CoV-2 transmission details). Although the study provided crucial findings about the misinformation surrounding COVID-19, the use of predefined categories might have biased the

analysis towards preexisting knowledge and limited the discovery of nuanced or unexpected themes.

4.4.2. Implications

This research illustrated the value of social media monitoring for obtaining insights into the knowledge, attitudes, and opinions of people across the world regarding public health issues. By identifying the various user groups involved in conversations around diphtheria on X, this study offered a foundation for the development of tailored communication strategies that address the specific concerns and informational needs of different user groups.

The analysis brought to light the need for governments and public health institutions to increase their participation in social media conversations relating to diphtheria and other healthcare issues. The importance of these entities actively engaging in social media is exemplified by a 2022 study by Zhai *et al*. [386]. By surveying 226 residents of the city of Wuhan in China during the first three weeks of COVID-19 lockdown, they found that information shared by government entities on the social media platforms Weibo, WeChat, and TikTok positively correlated with people's trust in local governments and knowledge of the disease transmission routes, symptoms, and preventive measures.

Public health institutions and practitioners should also partner with social media platforms and popular accounts (e.g., those of celebrities) to amplify accurate health-related information. For instance, from October 2018 to March 2019, a total of 117 social media influencers (i.e.,

individuals with 500 to 10,000 followers on at least one social media account) participated in a digital campaign to deliver positive flu vaccine-related information in the US, particularly among African American and Hispanic communities [387]. The campaign generated nearly 70,000 engagements (likes, shares, or comments), reaching a potential audience of 9.9 million individuals. Cross-sectional pre- and post-campaign surveys showed significant increases in knowledge and positive perceptions regarding the flu vaccine among respondents exposed to the campaign compared to the control group.

This study has also highlighted the necessity for platforms like X to continue to refine their systems for the detection and removal of false or misleading information, while guiding users on how to report such content. This is crucial because, as demonstrated in a 2018 study that analyzed 126,000 rumors and news cascades shared on X by around 3 million people from 2006 to 2017 [388], false news spreads significantly farther, faster, deeper, and more broadly than true news.

Finally, the themes and subthemes emerging from this study can serve as a framework for further investigation into the factors contributing to vaccine hesitancy. Future research on this subject should strive to incorporate posts in multiple languages and from different social media platforms to ensure the inclusion of a wider range of perspectives.

4.4.3. Limitations and strengths

This study had some limitations. Although a relatively extensive dataset was used, due to X's API restrictions, only a 1% random sample of all posts relating to diphtheria published during the study period were

collected. Therefore, it is uncertain whether included posts fully captured X users' views on diphtheria. Nevertheless, the recurrent appearance of specific key themes during this analysis, alongside the absence of new patterns, suggested that saturation may have been reached.

Only posts in English were examined, which resulted in the perspectives of non-English speaking populations not being fully captured, thereby limiting the generalizability of this study's findings. As X posts have less than 280 characters (with a limit of 140 characters until 2017) [389], it was sometimes difficult to understand the context and meaning of these messages, making it challenging to interpret and analyze the conversations. The collected data were also noisy and contained several unintelligible posts, which were excluded during data preprocessing using both automated and manual methods to mitigate the impact of these posts on the overall analysis. Despite these efforts, it is still possible that some relevant information was excluded, which may have affected the breadth and depth of the research findings.

Given that this research was based solely on data collected from X, the study's conclusions cannot be applied to other social media platforms because of the different behaviours and demographic profiles of their users. Due to their different designs and features, X is primarily used for real-time information dissemination, obtaining and sharing news updates, and engaging in public conversations, whereas Facebook is mainly used for maintaining personal connections and community building [390, 391]. Past research showed that X users are generally younger, wealthier, and better educated than users of other social media platforms like Facebook [392, 393]. These variations may account for

some of the divergent beliefs held by users of the two platforms. For instance, in a 2018 study, using a survey administered to over 2,500 people worldwide, Benoit and Mauldin [394] found that X users were more knowledgeable about vaccination and had more positive attitudes towards vaccines compared to Facebook users.

Lastly, as mentioned in Section 4.2.10, it is crucial to acknowledge that the conduct of this social media analysis and the interpretation of its data may have been influenced by our backgrounds, cultures, training, and prior diphtheria knowledge [351-353]. These are likely to have been different from those of many X users whose posts we analyzed. Being transparent about this limitation is paramount to retaining the overall credibility of this study's findings. To achieve richer and more rounded insights, future inquiries on the public discourse surrounding diphtheria could benefit from incorporating the voices and expertise of researchers from different cultural perspectives, particularly those living in high-burden settings.

Despite these limitations, this study presented several methodological strengths. The integration of quantitative and qualitative techniques provided a multi-dimensional lens that enhanced the depth and validity of the research findings. By employing advanced analytical techniques, it was possible to process and analyze large volumes of data, uncovering subtle patterns and trends in diphtheria-related conversations. Grounded Theory allowed to capture the complexity and diversity of public perceptions and attitudes towards the disease, which enhanced the overall robustness of the study. The use of an extensive dataset spanning ten years and covering the whole world enabled the detection of geographical areas of intense social media interest on diphtheria;

identification of recurrent themes and concerns around the disease; and exploration of shifts in global sentiment and discourse in response to various external events like the COVID-19 pandemic, outbreaks of other vaccine-preventable diseases, and the introduction of new vaccination policies. Overall, this study offered a multifaceted perspective on the online discourse surrounding diphtheria, providing valuable insights into people's engagement with health information on social media.

4.4.4. Conclusion

To the best of current knowledge, this is the first study that investigated the global discourse surrounding diphtheria on any social media platform. This research could serve as a model for exploring people's knowledge, attitudes, and opinions regarding public health topics in the digital age. Using over a decade worth of data and the integration of both quantitative and qualitative methods, this study demonstrated that various voices engaged in online discussions about the disease, resulting in the emergence of multiple themes, with vaccination being the main subject of interest. Crucially, the study highlighted the need for health authorities to increase their digital presence, improve user engagement, and implement partnerships to amplify accurate information, whereas social media platforms should continue to refine their systems for detecting misinformation. While some of these recommendations have already been proposed in the past, the current threat posed by vaccine-preventable diseases like diphtheria make their implementation more critical than ever.

Chapter 5. Discussion

5.1. Key findings

5.1.1. Main insights from the research studies

This thesis has presented an in-depth exploration of the global persistence of diphtheria through three separate but interconnected studies: a systematic review and meta-analysis, a GIS and spatial analysis, and a mixed-methods analysis of social media discourse. This approach yielded complementary insights, which together contribute to a more holistic understanding of the factors underlying diphtheria persistence.

The systematic review and meta-analysis identified multiple risk factors for diphtheria. Incomplete vaccination emerged as the most critical risk factor, with moderate quality evidence underpinning this finding. This observation is supported by extensive research on vaccine-preventable diseases [202-205], which demonstrated that individuals who do not complete their vaccination schedules are more susceptible to infections. Additional identified risk factors included contact with a person with skin lesions, and low knowledge of diphtheria. Further factors potentially linked to the disease were poor hygiene practices, such as sharing utensils and infrequent bathing. This aligns with past studies showing that good hygiene can reduce the risk of respiratory illnesses [223, 224]. The evidence quality for all these other factors was generally low.

The spatial analysis uncovered significant disparities in the reported diphtheria case rate across Haiti's departments and communes between December 2014 and June 2021. The analysis also revealed a positive association between this rate and healthcare facility density, suggesting that increased access to health services facilitated the detection of diphtheria cases [309, 310]. Furthermore, the study showed that the reported case rate was positively associated with the degree of urbanization. This could be due to urban areas being typically marked by crowded living conditions, alongside high population movement and social mixing, which collectively can facilitate disease spread [315, 316]. Additionally, the rate was negatively associated with female literacy, indicating that educated women may be more likely to understand health messages and the importance of protective measures like vaccination and personal hygiene, thereby lowering the risk of disease for themselves and their children [227, 228].

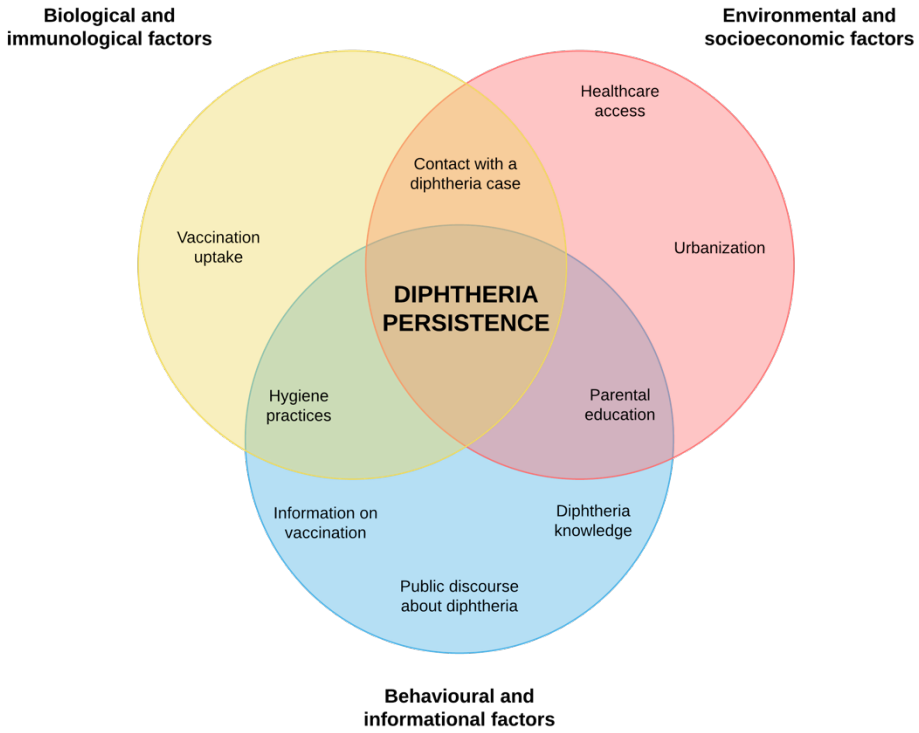
The examination of diphtheria-related posts published on X from 2012 to 2022 provided unique insights into the complex and sometimes contentious discourse surrounding the disease. The study revealed that the platform gave voice to a wide range of narratives and themes. Interestingly, diphtheria was frequently conflated with other vaccine-preventable diseases, suggesting a potential propensity of users to create connections between issues perceived as similar. News alerts and personal stories often revealed the devastating consequences of diphtheria, humanizing the statistical data on the disease. Expressions of gratitude towards vaccines and healthcare providers highlighted the beneficial impact of vaccination and other public health interventions. These positive testimonials were offset by posts coming from a small but vocal minority that propagated misinformation and conspiracy theories,

thereby attempting to erode public confidence in health institutions – as seen in previous studies of social media discussions [379, 380, 385]. The relatively low proportion of such content in this analysis may partly be attributable to X's reinforced regulations against misinformation following the COVID-19 pandemic – an event that effectively accelerated the polarization of public opinion on matters of health and science [380, 384, 395].

5.1.2. A theoretical model of diphtheria persistence

The synthesis of the insights from the three studies allowed the development of a theoretical model for understanding the determinants of diphtheria persistence (**Figure 5.1**). By illustrating how biological, immunological, environmental, socioeconomic, behavioural, and informational factors interact to sustain the disease, this model highlights the complexity of diphtheria dynamics.

Figure 5.1. Theoretical model of diphtheria persistence



The cornerstone of diphtheria persistence is vaccination uptake. Without full immunization, including booster doses, individuals remain vulnerable to the disease [60, 61, 65]. This biological susceptibility contributes to the emergence and sustenance of diphtheria outbreaks, since gaps in herd immunity facilitate the spread of the disease within a given population [58].

Both vaccination coverage and diphtheria persistence are significantly influenced by environmental and socioeconomic factors. First among these factors is the access to healthcare services. In areas with a low density of healthcare facilities, achieving optimal vaccination coverage is more challenging due to accessibility issues and insufficient resources, which can lead to high rates of incomplete vaccination, increasing the risk of outbreaks [82, 92-94, 309, 310]. Areas with low healthcare density are less likely to detect and report all diphtheria cases, which negatively impacts the initiation of treatment, thereby increasing the severity of disease and lowering the likelihood of survival [60, 68, 99]. Moreover, the underreporting or late reporting of cases also affects the capacity of public health authorities to isolate infected individuals and implement other necessary interventions, enabling the propagation of the infection [60, 68, 99].

The degree of urbanization introduces an additional layer of complexity. Due to their high population movement, significant social mixing, and often crowded living conditions, urban areas provide ample opportunities for people to come into contact with diphtheria cases, thus increasing the risk of contracting the disease, particularly if they are not fully vaccinated [315, 316].

Another socioeconomic factor that impacts diphtheria persistence is low parental education and literacy. Less educated parents may experience communication difficulties, making it more challenging for them to effectively engage with the healthcare system to get their children vaccinated [227, 228, 230]. Furthermore, less educated parents may reside in areas at high risk of diphtheria due to the inherent inadequate healthcare, sanitation, and vaccination [227, 231, 232]. Finally, less educated parents may possess lower literacy skills, reducing their receptiveness to information related to diphtheria, vaccination, and other preventive practices [227-229]. This, in turn, increases the likelihood of them engaging in poor hygiene practices, such as sharing household items and infrequent bathing, which heightens their susceptibility to diphtheria and other respiratory infections [223, 224].

Knowledge, attitudes, and behaviours towards diphtheria, vaccination, and other preventive health measures are influenced by the public perceptions and discourse around these subjects [82-84]. Misinformation about diphtheria and vaccines, as seen on social media platforms like X, can result in people delaying or completely avoiding vaccination due to confusion and fear [396, 397]. Ultimately, this contributes to creating areas with high concentrations of unimmunized individuals who contribute to the persistence of diphtheria.

5.2. Implications of findings

5.2.1. Implications for public health practice

Through the analysis, synthesis, and interpretation of various data, this research generated actionable insights. The systematic review's meta-analysis produced clear summary risk estimates, which could guide public health interventions and policies relating to diphtheria. The

spatial analysis revealed key disease trends and patterns that could inform response strategies in Haiti and other similar contexts. The social media analysis identified emerging themes and patterns in the public discourse surrounding diphtheria, providing a basis for the development of effective communication strategies.

The multiplicity, diversity, and interconnectedness of factors contributing to the persistence of diphtheria, which were uncovered by this research, emphasize the need for comprehensive and integrated public health strategies to effectively combat and prevent the disease. These strategies should seek to raise vaccination coverage, scale up access to healthcare services, strengthen surveillance systems, increase health literacy, and improve sanitation and hygiene conditions.

Incomplete vaccination presented the strongest association with diphtheria in the systematic review. Therefore, it is evident that vaccination programmes should focus on increasing coverage for the primary series and booster doses, particularly in areas at high risk of diphtheria due to suboptimal immunization levels, socioeconomic conditions, healthcare availability, and sanitary infrastructures. Mass vaccination campaigns like those conducted during the 1990s in the former Soviet Union are testament to the effectiveness of these interventions to control diphtheria outbreaks [235, 398]. Countries with limited resources can consider adopting innovative alternatives such as school-based vaccinations, which have been successful in several LMICs [106, 107]. These interventions will require overcoming logistical hurdles and expanding access to vaccines by, for instance, increasing their production and ensuring their availability at healthcare facilities [82, 399]. Another way to increase access to vaccines and amplify public

health impact include integrating vaccination with other healthcare services [400, 401].

As argued in the spatial analysis, diphtheria is likely underreported due to several factors. This highlights the need to strengthen surveillance systems to facilitate the rapid detection of diphtheria outbreaks, tracking of cases and contacts in real-time, and implementation of public health interventions. Enhanced surveillance requires establishing standardized protocols for case finding and contact tracing, reinforcing laboratory capacities, and improving data collection and reporting mechanisms [60, 68, 73]. This involves ensuring the availability of adequate testing equipment, alongside training healthcare professionals [76, 77]. Moreover, public health authorities should consider integrating spatial analysis tools into surveillance frameworks to identify emerging hotspots and obtain a more comprehensive picture of disease dynamics, which can ultimately guide the strategic allocation of resources to high-risk areas [16, 17].

Findings from the spatial analysis suggest that the presence of clinics and hospitals can bolster the early identification of diphtheria cases. This implies that healthcare services should be scaled up, for example through the construction and renovation of hospitals, clinics, and laboratories [402, 403], alongside the use of mobile clinics [404, 405]. In some settings, telemedicine and mobile health technologies could be leveraged to bring healthcare services closer to patients by enabling remote consultations, diagnoses, and monitoring, thereby reducing the need to travel long distances [406, 407]. Furthermore, public-private partnerships should be explored to mobilize additional resources and expertise, which would contribute to expanding

healthcare delivery, infrastructure development, and medical advancements [408, 409]. Additionally, policies aimed at achieving universal health coverage should continue to be developed and implemented to help households reduce financial expenditures associated with diphtheria, which will likely encourage patients to seek medical care early, leading to more timely diagnosis and treatment of diphtheria [410, 411].

The social media analysis revealed posts alerting on shortages of essential medicines in settings with ongoing diphtheria outbreaks. Efficient supply chain systems should thus be established to ensure the continuous availability of medical supplies for the treatment of identified cases [412, 413].

Given that inadequate knowledge about diphtheria emerged as a potential risk factor in the systematic review and that the social media analysis found misinformation about the disease in posts published on X, it would be important for public health programmes to integrate educational interventions. These should aim to raise awareness about diphtheria and preventive measures, encourage good hygiene practices (e.g., handwashing, surface disinfection, adequate indoor ventilation), and improve overall health literacy [414, 415]. These initiatives should be designed to meet the unique requirements of different population groups, particularly those with low educational levels and that live in socioeconomically disadvantaged areas. Thanks to their extensive reach, schools and community centres can serve as hubs for these initiatives [414, 415]. Community health workers, which have become integral part of healthcare services in many LMICs, can also play an instrumental role

in these efforts due to their connections and deep understanding of local realities [416, 417].

To connect with the public and disseminate accurate health information, health authorities should also harness the potential of social media platforms [42, 371, 418]. The continuous monitoring of online discourse facilitates the identification of emerging public concerns, enabling the development of targeted communication strategies that can inform the public and dispel myths [46-48]. Collaborating with trusted voices like influencers and community leaders can significantly amplify people's receptiveness to health messages, thereby increasing their impact [387]. Additionally, health authorities should partner with social media platforms to refine their algorithms for the detection and mitigation of misinformation.

Since poor hygiene may contribute to acquiring diphtheria as hinted by the systematic review, and to create a lasting impact, the above-mentioned interventions should be implemented alongside broader sanitation and hygiene policies. Examples of these policies include constructing new water treatment facilities, modernizing existing infrastructures, and expanding water distribution networks to underserved communities [419, 420]. Guaranteeing access to adequate public sanitation facilities, including improved sewage systems, is also critical. Offering subsidies for sanitation improvements could incentivize households to invest in higher-quality facilities [421]. Finally, there is a need to enact and enforce water quality and sanitation regulations to ensure that service providers meet essential health and safety standards [419, 420].

The proposed interventions are key for controlling diphtheria. Nevertheless, their full implementation could be challenging for many LMICs. Haiti vividly illustrates the obstacles present in such settings, where progress in public health has often been hampered by socioeconomic instability, infrastructure breakdown, and humanitarian emergencies. Therefore, the recommended interventions should be adapted to local realities. For instance, affected countries could consider a step-by-step approach that focuses on urgent needs first while building capacity over time. In the case of Haiti, initial efforts could concentrate on reinforcing surveillance to accurately assess the disease burden and identify priority areas for interventions. In parallel, emergency vaccination and educational campaigns could be conducted in high-risk areas, while supply chains for vaccines and essential medicines are strengthened.

5.2.2. Implications for public health research

This research highlighted a paucity of studies on diphtheria, particularly from areas that have accounted for most of the disease burden in recent years, such as the African region. There is a clear need to increase R&D investments towards neglected diseases such as diphtheria in order to deepen the knowledge base to inform and optimize prevention and control efforts against the disease. Addressing this gap is particularly critical given COVID-19's lingering effects on routine vaccinations, with millions of children still lacking protection against diphtheria [249, 250].

Findings from this thesis revealed key areas for further investigation, which can enhance the comprehension of the epidemiology of diphtheria, ultimately strengthening public health interventions to tackle the disease. Future research on diphtheria should incorporate

mixed-methods approaches, longitudinal studies, and comparative analyses.

For instance, mixed-methods approaches that integrate spatial analyses of diphtheria cases and interviews with local health professionals, community leaders, and residents (as originally planned for this research project) could be used to gain a better understanding of the spatiotemporal patterns of the disease, the factors contributing to the observed trends, and the specific prevention and response challenges. The value and feasibility of combining interviews and spatial analyses has been demonstrated by previous studies that investigated tuberculosis in Madagascar [422] and self-harm in the UK [423]. Collaborations with microbiologists would enable the isolation and sequencing of *C. diphtheriae* strains to obtain deeper insights into the transmission patterns of the disease, uncover outbreak sources, and monitor trends in antibiotic resistance, which could not be explored through this thesis [424, 425]. Furthermore, studies that incorporate epidemiological and economic analyses would allow researchers to compare the cost-effectiveness of interventions like those proposed in Section 5.2.1.

Longitudinal studies can help capture evolving trends and shifts in diphtheria transmission and disease burden [426]. For example, such studies could examine the duration of immunity from infection or vaccination in young children, given the absence of long-term follow-up studies targeting this specific demographic group [427], to determine whether current vaccination guidelines need to be adjusted to ensure adequate protection. By following individuals who contract diphtheria over time, longitudinal studies could provide additional information

regarding the range and severity of symptoms of diphtheria, alongside its long-term sequelae, providing a comprehensive picture of the health impact on individuals who acquire the disease.

Comparative studies could help uncover how environmental, socioeconomic, and behavioural differences across countries may influence the epidemiology of diphtheria [428]. Examining data from various settings would enable the identification of common patterns in the disease's occurrence, recurring hurdles in control efforts, and effective public health strategies. For instance, research comparing response strategies in HICs and LMICs or contrasting LMICs that have achieved different levels of success in reducing diphtheria rates could pinpoint impactful interventions and explore how these can be adapted to diverse locations. Such comparative analyses could also expose weaknesses in current responses, while identifying key contextual factors that facilitate or hinder the control of the disease in different settings. Furthermore, comparative studies could also explore how different populations seek and access medical care for diphtheria, guiding policies to enhance early case detection and treatment.

The rising influence of social media on public perceptions and behaviors related to diphtheria and other vaccine-preventable diseases calls for further research in this area [40, 42]. Future studies could use findings from this thesis to further investigate the social media discourse surrounding public health issues, vaccine hesitancy, and the role of misinformation. For example, researchers could evaluate the effectiveness of different messaging strategies on social media platforms to identify the most effective methods for increasing vaccine acceptance and reducing the spread of false information. Future studies could use

social media data to track in real time public interest and concerns relating to diphtheria outbreaks, which can then inform risk communication strategies.

While social media analysis offers a valuable window into the public discourse surrounding specific issues, it does not capture the entirety of people's knowledge and perceptions, especially in LMICs where internet access and social media usage may be limited. Therefore, future research should also consider other methods such as focus groups, interviews, surveys, and the analysis of information coming from traditional media (e.g., press, television, radio) to gain a deeper understanding of public attitudes and opinions [429, 430].

Lastly, given the increasing number of reported diphtheria cases, particularly in LMICs as highlighted through this thesis, urgent research is needed for the development of new diagnostic tools and treatments for the disease [431, 432]. Innovative diagnostic techniques could offer more rapid and accurate detection of diphtheria, improving surveillance and outbreak control efforts. Novel therapeutic options for diphtheria could reduce the overall disease burden and dramatically improve clinical outcomes, particularly for severe cases and vulnerable populations like infants, the elderly, and the immunocompromised. It will be imperative to validate these new technologies in different settings to ensure their effectiveness, while guaranteeing their accessibility in both well-resourced and under-resourced environments.

5.3. Limitations and strengths

5.3.1. Limitations

The findings and implications from this research should be interpreted in light of its limitations. While the limitations specific to each individual study have been discussed in their respective chapters, there are also limitations that apply to all three studies. These broader limitations reflect the inherent challenges of undertaking rigorous, in-depth epidemiological investigations, particularly in resource-constrained settings.

One of the most significant limitations across the three studies was the dependence on observational data, which by nature are susceptible to potential biases, such as recall bias, selection bias, and confounding [150, 153, 433]. For instance, although the systematic review identified potential risk factors for diphtheria, it relied on evidence that was mostly rated as low or very low. This affected the robustness of the study's conclusions, making it impossible to establish definitive causal relationships between diphtheria and the risk factors. Similarly, despite providing valuable insights regarding the epidemiology of the disease in Haiti, the spatial analysis could not irrefutably link specific risk factors to the outbreak.

Another common constraint concerned the completeness of the data. The systematic review identified risk factors from a limited number of studies of varying methodological rigour. Likewise, the results of the spatial analysis were likely biased by the underreporting of diphtheria cases in Haiti. The analysis of social media posts was also affected by the relatively small sample size, the use of a single platform,

and the sole inclusion of posts in English, which restricted the generalizability and comprehensiveness of the research findings.

The reliance on manual processes introduced some degree of subjectivity [434, 435]. The systematic review depended on human judgment in the screening and inclusion of papers, extraction of data, and quality assessments, which likely added a certain level of uncertainty to the findings. Similarly, in the social media analysis, X posts were manually excluded if considered unintelligible, which may have led to relevant information being removed inadvertently. Furthermore, the research team members' backgrounds and prior knowledge about diphtheria, including their respective biases, inevitably influenced the formulation of the studies and the interpretation of findings.

Finally, a key limitation of this thesis was the inability to comprehensively answer the central research question of why diphtheria persists in Haiti. Despite providing crucial insights into the epidemiology and factors shaping reported diphtheria case rates in the country, the spatial analysis did not fully elucidate the dynamics and complexity of the outbreak. A major contributing element to this limitation was Haiti's contextual situation, characterized by social unrest, economic hardship, and a deteriorating healthcare system, which not only exacerbated the transmission of the disease but also hindered holistic research efforts. The lack of primary qualitative data from Haiti due to the challenges of conducting field research in the country was a critical gap, precluding a nuanced understanding of local perspectives and experiences relating to the disease.

5.3.2. Strengths

Despite the above-mentioned limitations, this thesis also exhibited significant methodological strengths that underscored the depth, robustness, and credibility of the research findings, while contributing to the body of knowledge on diphtheria persistence.

One of the most notable strengths was the comprehensive nature of the adopted research approach. Each study employed extensive data collection methods. By using a broad search strategy with no restrictions on dates, languages, or geography, the systematic review captured studies involving thousands of participants from nearly all continents and spanning several decades, generating global insights on diphtheria risk factors, which strengthened the generalizability of the findings. Similarly, the spatial analysis leveraged a subnational dataset of diphtheria cases in Haiti that covered over six years, providing a granular exploration of temporal and spatial trends of the disease in the country. Furthermore, the use of both official and open-source data enhanced the thoroughness of the analysis, ensuring a holistic overview of diphtheria in Haiti. The social media analysis also had an expansive methodological approach characterized by the examination of a decade's worth of data from around the world, offering a unique window into the public discourse surrounding diphtheria.

The three studies also benefitted from methodological rigor and the use of various analytical methods, including innovative tools. In the systematic review, only studies with similar exposure definitions were included in the meta-analysis, increasing the reliability of the generated summary risk estimates. The spatial analysis employed both descriptive and advanced GIS techniques, revealing diphtheria hotspots and factors

associated with the reported case rate in Haiti. Instead, the social media analysis combined quantitative and qualitative methods to identify measurable patterns in diphtheria-related conversations, uncover emerging themes and subthemes, and highlight how social media contributes to shaping public perceptions and behaviors.

The interdisciplinary nature of this research allowed the triangulation of data available from various sources and disciplines, enabling the identification of different factors that contribute to diphtheria transmission. Such approach not only enhanced the validity of the research findings but also facilitated the development of a theoretical model of diphtheria persistence, offering a holistic and nuanced understanding of the disease's intricate nature. While the model was not exclusively derived from Haitian data, it provided insights that may still be highly relevant to the country and other similar settings. In turn, these insights could guide the development of more effective public health interventions, even in the face of major data and operational constraints.

Lastly, my professional experience with WHO in Haiti (and afterwards in Nigeria) provided me with a direct exposure to some of the contextual elements influencing the control and prevention of the disease. My close connection to the topic may have led to preconceived views about the underlying causes of diphtheria persistence, potentially influencing the interpretation of the findings. Nevertheless, this real-world experience was crucial for identifying key risk factors and public discourse themes, as well as understanding the epidemiological patterns of the disease in Haiti. Furthermore, this practical knowledge ensured that the recommendations from this research reflected findings from the three

studies along with firsthand insights from actual responses to diphtheria outbreaks.

5.4. Conclusion

This thesis shed light on the global persistence of diphtheria, integrating insights from a systematic review, a spatial analysis, and a mixed-methods examination of social media discourse. The adopted methodological approach, which incorporated various analytical techniques and data sources, enhanced the trustworthiness and validity of the research findings. This thesis contributes to the existing body of knowledge on diphtheria by illustrating how biological, immunological, environmental, socioeconomic, behavioural, and informational factors potentially interact to sustain the disease. Given the complexity of diphtheria epidemiology, health authorities should strive to implement multifaceted prevention and control strategies that address these multiple factors. While remaining context-specific, these strategies should all aim to raise vaccination coverage, scale up access to healthcare services, strengthen surveillance systems, increase health literacy, and improve sanitation and hygiene conditions. These strategies should be accompanied by renewed investments in diphtheria research not only to enhance comprehension of the disease but also to guide the development and implementation of effective public health interventions, especially in countries most vulnerable to diphtheria like Haiti. Future research should continue to foster interdisciplinary collaborations and include the use of longitudinal studies, comparative analyses, and innovative tools, such as GIS and social media monitoring. The current worldwide resurgence of diphtheria and other vaccine-preventable diseases underscores the urgent need for action.

Bibliography

1. Smallman-Raynor, M.R., et al., *Pandemics, I: Pandemics in History*. 2022, United Kingdom: Oxford University Press, Incorporated: United Kingdom.
2. Snowden, F.M., *Epidemics and Society : From the Black Death to the Present*. 1st ed. 2019: New Haven : Yale University Press.
3. Littman, R.J., *The plague of Athens: epidemiology and paleopathology.* Mt Sinai J Med, 2009. **76**(5): p. 456-67.
4. DeWitte, S.N., *Mortality Risk and Survival in the Aftermath of the Medieval Black Death.* PLOS ONE, 2014. **9**(5): p. e96513.
5. Spreeuwenberg, P., M. Kroneman, and J. Paget, *Reassessing the Global Mortality Burden of the 1918 Influenza Pandemic.* American Journal of Epidemiology, 2018. **187**(12): p. 2561-2567.
6. World Health Organization, *COVID-19 epidemiological update, edition 162, 22 December 2023*. 2023, World Health Organization,: Geneva.
7. M'Ikanatha, N.M., *Infectious Disease Surveillance*. 2nd ed, ed. R. Lynfield, C.A. Van Beneden, and H. de Valk. 2013: Newark : John Wiley & Sons, Incorporated.
8. Gaynes, R.P., *Germ Theory: Medical Pioneers in Infectious Diseases*. 2023: Wiley.
9. Liu, Q., et al., *Advances in the application of molecular diagnostic techniques for the detection of infectious disease pathogens (Review).* Molecular medicine reports, 2023. **27**(5).
10. Nolan, T., *PCR Technology : Current Innovations, Third Edition*. 3rd ed, ed. S.A. Bustin. 2013: Baton Rouge : Taylor & Francis Group.

11. Voelkerding, K.V., S.A. Dames, and J.D. Durtschi, *Next-Generation Sequencing: From Basic Research to Diagnostics.* Clinical chemistry (Baltimore, Md.), 2009. **55**(4): p. 641-658.
12. Lu, R., et al., *Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding.* The Lancet (British edition), 2020. **395**(10224): p. 565-574.
13. Zhu, N., et al., *A Novel Coronavirus from Patients with Pneumonia in China, 2019.* The New England journal of medicine, 2020. **382**(8): p. 727-733.
14. Kumari, M., et al., *A critical overview of current progress for COVID-19: development of vaccines, antiviral drugs, and therapeutic antibodies.* Journal of Biomedical Science, 2022. **29**(1): p. 68.
15. Rabaan, A.A., et al., *SARS-CoV-2/COVID-19 and advances in developing potential therapeutics and vaccines to counter this emerging pandemic.* Annals of Clinical Microbiology and Antimicrobials, 2020. **19**(1): p. 40.
16. Smith, C.M., et al., *Spatial methods for infectious disease outbreak investigations: systematic literature review.* Eurosurveillance, 2015. **20**(39): p. 6-26.
17. Kirby, R.S., E. Delmelle, and J.M. Eberth, *Advances in spatial epidemiology and geographic information systems.* Annals of Epidemiology, 2017. **27**(1): p. 1-9.
18. Backer, J.A. and J. Wallinga, *Spatiotemporal Analysis of the 2014 Ebola Epidemic in West Africa.* PLoS Comput Biol, 2016. **12**(12): p. e1005210.
19. Carroll, M.W., et al., *Temporal and spatial analysis of the 2014-2015 Ebola virus outbreak in West Africa.* Nature, 2015. **524**(7563): p. 97-101.

20. Lau, M.S., et al., *Spatial and temporal dynamics of superspreading events in the 2014-2015 West Africa Ebola epidemic.* Proc Natl Acad Sci U S A, 2017. **114**(9): p. 2337-2342.
21. Suchar, V.A., et al., *An Exploration of the Spatiotemporal and Demographic Patterns of Ebola Virus Disease Epidemic in West Africa Using Open Access Data Sources.* Appl Geogr, 2018. **90**: p. 272-281.
22. Allan, M., et al., *High-resolution spatial analysis of cholera patients reported in Artibonite department, Haiti in 2010-2011.* Epidemics, 2016. **14**: p. 1-10.
23. Blackburn, J.K., et al., *Household-level spatiotemporal patterns of incidence of cholera, Haiti, 2011.* Emerg Infect Dis, 2014. **20**(9): p. 1516-9.
24. Griffiths, K., et al., *Delineating and Analyzing Locality-Level Determinants of Cholera, Haiti.* Emerg Infect Dis, 2021. **27**(1): p. 170-181.
25. Gao, J., et al., *Evidence-driven spatiotemporal COVID-19 hospitalization prediction with Ising dynamics.* Nat Commun, 2023. **14**(1): p. 3093.
26. Mohammadi, A., et al., *Measuring COVID-19 vaccination coverage: an enhanced age-adjusted two-step floating catchment area model.* Infect Dis Poverty, 2021. **10**(1): p. 118.
27. Tiu, A., et al., *Characterizing the Spatiotemporal Heterogeneity of the COVID-19 Vaccination Landscape.* Am J Epidemiol, 2022. **191**(10): p. 1792-1802.
28. Wang, P., X. Zheng, and H. Liu, *Simulation and forecasting models of COVID-19 taking into account spatio-temporal dynamic characteristics: A review.* Front Public Health, 2022. **10**: p. 1033432.

29. Watson, S.I., et al., *Spatiotemporal analysis of the first wave of COVID-19 hospitalisations in Birmingham, UK.* BMJ Open, 2021. **11**(10): p. e050574.

30. Xu, Z. and B. Jiang, *Effects of Social Vulnerability and Spatial Accessibility on COVID-19 Vaccination Coverage: A Census-Tract Level Study in Milwaukee County, USA.* Int J Environ Res Public Health, 2022. **19**(19).

31. Afolabi, M.O., et al., *Prevalence and distribution pattern of malaria and soil-transmitted helminth co-endemicity in sub-Saharan Africa, 2000-2018: A geospatial analysis.* PLoS Negl Trop Dis, 2022. **16**(9): p. e0010321.

32. Bhatt, S., et al., *The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015.* Nature, 2015. **526**(7572): p. 207-211.

33. Diouf, I., et al., *Climate Variability and Malaria over West Africa.* Am J Trop Med Hyg, 2020. **102**(5): p. 1037-1047.

34. Giardina, F., et al., *Effects of vector-control interventions on changes in risk of malaria parasitaemia in sub-Saharan Africa: a spatial and temporal analysis.* Lancet Glob Health, 2014. **2**(10): p. e601-15.

35. Kienberger, S. and M. Hagenlocher, *Spatial-explicit modeling of social vulnerability to malaria in East Africa.* Int J Health Geogr, 2014. **13**: p. 29.

36. Noor, A.M., et al., *The changing risk of Plasmodium falciparum malaria infection in Africa: 2000-10: a spatial and temporal analysis of transmission intensity.* Lancet, 2014. **383**(9930): p. 1739-47.

37. Jacquez, G.M., *Current practices in the spatial analysis of cancer: flies in the ointment.* Int J Health Geogr, 2004. **3**(1): p. 22.

38. Fotheringham, A.S., *Geographically weighted regression : the analysis of spatially varying relationships / A. Stewart Fotherington, Chris*

Brunsdon, Martin Charlton, ed. C. Brunsdon and M. Charlton. 2002, Chichester: Chichester : Wiley.

39. Boscoe, F.P., M.H. Ward, and P. Reynolds, *Current practices in spatial analysis of cancer data: data characteristics and data sources for geographic studies of cancer.* International Journal of Health Geographics, 2004. **3**(1): p. 28.
40. Aiello, A.E., A. Renson, and P.N. Zivich, *Social Media- and Internet-Based Disease Surveillance for Public Health.* Annu Rev Public Health, 2020. **41**: p. 101-118.
41. Brownstein, J.S., C.C. Freifeld, and L.C. Madoff, *Digital disease detection--harnessing the Web for public health surveillance.* N Engl J Med, 2009. **360**(21): p. 2153-5, 2157.
42. Wang, A., et al. *A Review of Social Media Data Utilization for the Prediction of Disease Outbreaks and Understanding Public Perception*. Big Data and Cognitive Computing, 2023. **7**, DOI: 10.3390/bdcc7020072.
43. Golder, S., et al., *A chronological and geographical analysis of personal reports of COVID-19 on Twitter from the UK.* Digit Health, 2022. **8**: p. 20552076221097508.
44. Gharavi, E., N. Nazemi, and F. Dadgostari, *Early outbreak detection for proactive crisis management using twitter data: Covid-19 a case study in the us.* arXiv preprint arXiv:2005.00475, 2020.
45. Turiel, J., D. Fernandez-Reyes, and T. Aste, *Wisdom of crowds detects COVID-19 severity ahead of officially available data.* Scientific Reports, 2021. **11**(1): p. 13678.
46. Sinnenberg, L., et al., *Twitter as a Tool for Health Research: A Systematic Review.* Am J Public Health, 2017. **107**(1): p. e1-e8.
47. Salathé, M. and S. Khandelwal, *Assessing Vaccination Sentiments with Online Social Media: Implications for Infectious Disease*

Dynamics and Control. PLOS Computational Biology, 2011. **7**(10): p. e1002199.

48. Ng, Q.X., et al., *Public sentiment on the global outbreak of monkeypox: an unsupervised machine learning analysis of 352,182 twitter posts.* Public Health, 2022. **213**: p. 1-4.
49. Blank, G., *The digital divide among Twitter users and its implications for social research.* Social Science Computer Review, 2017. **35**(6): p. 679-697.
50. Charles-Smith, L.E., et al., *Using Social Media for Actionable Disease Surveillance and Outbreak Management: A Systematic Literature Review.* PloS one, 2015. **10**(10): p. e0139701-e0139701.
51. Sloan, L. and J. Morgan, *Who Tweets with Their Location? Understanding the Relationship between Demographic Characteristics and the Use of Geoservices and Geotagging on Twitter.* PLOS ONE, 2015. **10**(11): p. e0142209.
52. Sloan, L., et al., *Knowing the Tweeters: Deriving Sociologically Relevant Demographics from Twitter.* Sociological Research Online, 2013. **18**(3): p. 74-84.
53. GSMA Intelligence, *State of Mobile Internet Connectivity 2021*. 2021: London.
54. Group, I.P.C., *Global burden associated with 85 pathogens in 2019: a systematic analysis for the Global Burden of Disease Study 2019.* The Lancet infectious diseases, 2024.
55. Sykora, M., S. Elayan, and T.W. Jackson, *A qualitative analysis of sarcasm, irony and related #hashtags on Twitter.* Big Data & Society, 2020. **7**(2): p. 2053951720972735.
56. Wankhade, M., A.C.S. Rao, and C. Kulkarni, *A survey on sentiment analysis methods, applications, and challenges.* Artificial Intelligence Review, 2022. **55**(7): p. 5731-5780.

57. Wilson, A.E., et al., *Social media: A new tool for outbreak surveillance.* Antimicrobial Stewardship & Healthcare Epidemiology, 2021. **1**(1): p. e50.

58. World Health Organization, *Diphtheria vaccine: WHO position paper – August 2017.* Releve epidemiologique hebdomadaire, 2017. **92**(31): p. 417-435.

59. World Health Organization. *Diphtheria reported cases*. 2023 5 June 2023]; Available from: https://www.who.int/data/gho/data/indicators/indicator-details/GHO/diphtheria---number-of-reported-cases.

60. Heymann, D.L., *Control of communicable diseases manual / David L. Heymann, editor*. 20th ed. 2015, Washington, D.C.: Washington, D.C. : American Public Health Association.

61. Prod'hom, G. and J. Bille, *178 - Aerobic Gram-Positive Bacilli*, in *Infectious Diseases (4th ed.)*, J. Cohen, W.G. Powderly, and S.M. Opal, Editors. 2017, Elsevier. p. 1537-1552.e2.

62. Hamborsky, J., et al., *Epidemiology and prevention of vaccine-preventable diseases*. 13th ed. 2015: Atlanta, GA.: Dept. of Health & Human Services, Public Health Service, Centers for Disease Control and Prevention.

63. Sharma, N.C., et al., *Diphtheria.* Nature Reviews Disease Primers, 2019. **5**(1): p. 81.

64. Sellers, T.F., *Corynebacterium diphtheriae - Gram stain*. 1973, Centers for Disease Control and Prevention,.

65. Tiwari, T.S.P. and M. Wharton, *19 - Diphtheria Toxoid*, in *Plotkin's Vaccines*, S.A. Plotkin, et al., Editors. 2018, Elsevier. p. 261-275.e7.

66. World Health Organization, *The immunological basis for immunization series: module 2: diphtheria - Update 2009*. 2009, Geneva: World Health Organization.

67. Tiwari, T.S.P., *37 - Diphtheria*, in *Hunter's Tropical Medicine and Emerging Infectious Disease*, A.J. Magill, et al., Editors. 2013, W.B. Saunders: London. p. 402-406.
68. World Health Organization, *Vaccine Preventable Diseases Surveillance Standards*. 2018, World Health Organization: Geneva.
69. Truelove, S.A., et al., *Clinical and Epidemiological Aspects of Diphtheria:A Systematic Review and Pooled Analysis.* Clin Infect Dis, 2019.
70. Lal, S., *Dirty white pseudomembrane classically seen in diphtheria*. 2013, Wikimedia Commons.
71. Smith, D.H., *Diphtheria bull neck*. 1995, CDC Public Health Image Library.
72. Clark, W.A., *Corynebacterium diphtheriae - Gram stain*. n.d. , CDC Public Health Image Library.
73. World Health Organization, *WHO laboratory manual for the diagnosis of diphtheria and other related infections*. 2021, Geneva: World Health Organization.
74. Reading, N., *Corynebacterium diphtheriae*. 2011, Flickr.
75. Williams, M.M., et al., *Detection and Characterization of Diphtheria Toxin Gene-Bearing Corynebacterium Species through a New Real-Time PCR Assay.* J Clin Microbiol, 2020. **58**(10).
76. European Centre for Disease Prevention and Control, *Gap analysis on securing diphtheria diagnostic capacity and diphtheria antitoxin availability in the EU/EEA*. 2017, Stockholm: European Centre for Disease Prevention and Control,.
77. Gurung, S., et al., *Diphtheria diagnostic capacity in the Western Pacific Region.* Western Pac Surveill Response J, 2019. **10**(4): p. 46-49.

78. Acosta, A.M. and T.S. Tiwari, *KEY FEATURES.* Hunter's Tropical Medicine and Emerging Infectious Diseases E-Book, 2019: p. 439.
79. Kabanova, A. and R. Rappuoli, *CHAPTER 34 - Diphtheria*, in *Tropical Infectious Diseases: Principles, Pathogens and Practice (Third Edition)*, R.L. Guerrant, D.H. Walker, and P.F. Weller, Editors. 2011, W.B. Saunders: Edinburgh. p. 223-227.
80. UK Health Security Agency, *Public health control and management of diphtheria in England: 2023 guidelines*. 2023, UK Health Security Agency,: London, United Kingdom.
81. Will, R.C., et al., *Spatiotemporal persistence of multiple, diverse clades and toxins of Corynebacterium diphtheriae.* Nature Communications, 2021. **12**(1): p. 1500.
82. Phillips, D.E., et al., *Determinants of effective vaccine coverage in low and middle-income countries: a systematic review and interpretive synthesis.* BMC Health Serv Res, 2017. **17**(1): p. 681.
83. Smith, L.E., et al., *A systematic review of factors affecting vaccine uptake in young children.* Vaccine, 2017. **35**(45): p. 6059-6069.
84. Gudayu, T.W. and H.T. Mengistie, *COVID-19 vaccine acceptance in sub-Saharan African countries: A systematic review and meta-analysis.* Heliyon, 2023. **9**(2): p. e13037.
85. Novilla, M.L.B., et al., *Why Parents Say No to Having Their Children Vaccinated against Measles: A Systematic Review of the Social Determinants of Parental Perceptions on MMR Vaccine Hesitancy.* Vaccines (Basel), 2023. **11**(5).
86. Eggertson, L., *Lancet retracts 12-year-old article linking autism to MMR vaccines.* Cmaj, 2010. **182**(4): p. E199-200.
87. Gaia, A., *Social desirability bias and sensitive questions in surveys / by Alessandra Gaia*

edited by Paul Atkinson, Sara Delamont, Alexandru Cernat, Joseph W. Sakshaug & Richard A. Williams, ed. P. Atkinson, et al. 2020: London : SAGE Publications Ltd.

88. Lydon, P., et al., *Vaccine stockouts around the world: Are essential vaccines always available when needed?* Vaccine, 2017. **35**(17): p. 2121-2126.

89. Ezezika, O., et al., *What are the barriers and facilitators to polio vaccination and eradication programs? A systematic review.* PLOS Global Public Health, 2022. **2**(11): p. e0001283.

90. Guzman-Holst, A., et al., *Barriers to vaccination in Latin America: A systematic literature review.* Vaccine, 2020. **38**(3): p. 470-481.

91. Wilson, R.J., et al., *Understanding factors influencing vaccination acceptance during pregnancy globally: A literature review.* Vaccine, 2015. **33**(47): p. 6420-6429.

92. Albers, A.N., J. Thaker, and S.R. Newcomer, *Barriers to and facilitators of early childhood immunization in rural areas of the United States: A systematic review of the literature.* Preventive Medicine Reports, 2022. **27**: p. 101804.

93. Amponsah-Dacosta, E., B.M. Kagina, and J. Olivier, *Health systems constraints and facilitators of human papillomavirus immunization programmes in sub-Saharan Africa: a systematic review.* Health Policy and Planning, 2020. **35**(6): p. 701-717.

94. Bangura, J.B., et al., *Barriers to childhood immunization in sub-Saharan Africa: A systematic review.* BMC Public Health, 2020. **20**(1): p. 1108.

95. Lacombe-Duncan, A., P.A. Newman, and P. Baiden, *Human papillomavirus vaccine acceptability and decision-making among adolescent boys and parents: A meta-ethnography of qualitative studies.* Vaccine, 2018. **36**(19): p. 2545-2558.

96. Lutz, C.S., et al., *Understanding barriers and predictors of maternal immunization: Identifying gaps through an exploratory literature review.* Vaccine, 2018. **36**(49): p. 7445-7455.

97. Newman, P.A., et al., *Parents' uptake of human papillomavirus vaccines for their children: a systematic review and meta-analysis of observational studies.* BMJ Open, 2018. **8**(4): p. e019206.

98. Braveman, P., *The social determinants of health and health disparities / Paula Braveman*. 2023: New York, NY : Oxford University Press.

99. Centers for Disease Control and Prevention, *Manual for the surveillance of vaccine-preventable diseases*. 2008, Atlanta, GA.: Centers for Disease Control and Prevention.

100. World Health Organization, *WHO guide for standardization of economic evaluations of immunization programmes*. 2019, World Health Organization,: Geneva.

101. Griffiths, U., et al., *Costs of delivering COVID-19 vaccine in 92 AMC countries.* World Health Organization, Gavi, unicef. https://www. who. int/publications/m/item/costs-of-delivering-covid-19-vaccine-in-92-amc-countries, 2021.

102. Wang, X., et al., *Influenza vaccination strategies for 2020-21 in the context of COVID-19.* J Glob Health, 2020. **10**(2): p. 021102.

103. Sharp, A., et al., *High cholera vaccination coverage following emergency campaign in Haiti: Results from a cluster survey in three rural Communes in the South Department, 2017.* PLoS Negl Trop Dis, 2020. **14**(1): p. e0007967.

104. Khan, A.I., et al., *Feasibility, coverage and cost of oral cholera vaccination conducted by icddr,b using the existing national immunization service delivery mechanism in rural setting Keraniganj, Bangladesh.* Hum Vaccin Immunother, 2019. **15**(6): p. 1302-1309.

105. World Health Organization, *School vaccination readiness assessment tool*. 2013, World Health Organization,: Geneva.
106. Ladner, J., et al., *Assessment of eight HPV vaccination programs implemented in lowest income countries.* BMC Public Health, 2012. **12**(1): p. 370.
107. Paul, P. and A. Fabio, *Literature review of HPV vaccine delivery strategies: considerations for school- and non-school based immunization program.* Vaccine, 2014. **32**(3): p. 320-6.
108. Feldstein, L.R., et al., *School-based delivery of routinely recommended vaccines and opportunities to check vaccination status at school, a global summary, 2008–2017.* Vaccine, 2020. **38**(3): p. 680-689.
109. Gupta, P.S., et al., *Expanding COVID-19 vaccine access to underserved populations through implementation of mobile vaccination units.* Prev Med, 2022. **163**: p. 107226.
110. Palmer, K., et al., *COVID-19 vaccination roll-outs in eleven small countries within the WHO European region; Lessons learned.* Front Public Health, 2022. **10**: p. 959227.
111. Shukla, S., et al., *Optimizing vaccine distribution via mobile clinics: a case study on COVID-19 vaccine distribution to long-term care facilities.* Vaccine, 2022. **40**(5): p. 734-741.
112. Haddison, E.C., et al., *Successful polio supplementary immunisation activities in a security compromised zone - Experiences from the Southwest region of Cameroon.* Vaccine, 2018. **36**(46): p. 6961-6967.
113. Curry, D.W., et al., *Assessing the effectiveness of house-to-house visits on routine oral polio immunization completion and tracking of defaulters.* J Health Popul Nutr, 2014. **32**(2): p. 356-66.
114. Clarke, K.E.N., et al., *Global Epidemiology of Diphtheria, 2000-2017.* Emerging infectious diseases, 2019. **25**(10): p. 1834-1842.

115. Wagner, K.S., et al., *Diphtheria in the postepidemic period, Europe, 2000-2009.* Emerg Infect Dis, 2012. **18**(2): p. 217-25.

116. Mattos-Guaraldi, A.L., et al., *Diphtheria remains a threat to health in the developing world--an overview.* Mem Inst Oswaldo Cruz, 2003. **98**(8): p. 987-93.

117. Blumberg, L.H., et al., *The preventable tragedy of diphtheria in the 21st century.* Int J Infect Dis, 2018. **71**: p. 122-123.

118. Encyclopedia Britannica, *Haiti*, in *Encyclopedia Britannica,*. 2022, Encyclopedia Britannica,.

119. Pan American Health Organization / World Health Organization. *Canadian support in Haiti launches successful catch-up diphtheria vaccination campaign*. 2024 [cited 2024 20 July 2024]; Available from: https://www.paho.org/en/news/19-7-2024-canadian-support-haiti-launches-successful-catch-diphtheria-vaccination-campaign#:~:text=From%202014%20to%202023%2C%20there,against%20all%20vaccine%2Dpreventable%20diseases.

120. Pan American Health Organization / World Health Organization, *Epidemiological Update: Diphtheria in Hispaniola. 5 November 2021*. 2021: Washington, D.C.

121. Clerville, J., *Diphtheria Outbreak, Haiti, 2014-2017: An Epidemiological Profile and A Case Fatality Rate Trend Analysis.* International Journal of Infectious Diseases, 2018. **73**: p. 274-275.

122. Exavier, M.M., et al., *Diphtheria in Children in Northern Haiti.* J Trop Pediatr, 2018.

123. Charmaz, K., *Constructing grounded theory / Kathy Charmaz*. 2nd ed. ed. 2014, London: London : SAGE.

124. Glaser, G.B., L.A. Strauss, and L.E. Strutzel, *The Discovery of Grounded Theory*

Strategies for Qualitative Research. Nursing Research, 1968. **17**(4): p. 364-364.

125. Strauss, A.L., *Basics of qualitative research : techniques and procedures for developing grounded theory / Anselm Strauss, Juliet Corbin*. 2nd ed. ed, ed. J.M. Corbin. 1998, Thousand Oaks, Calif.

London: Thousand Oaks, Calif.

London : SAGE.

126. Creswell, J.W., *Qualitative inquiry & research design : choosing among five approaches / John W. Creswell, Cheryl N. Poth*. 4th ed., International Student ed. ed. Qualitative inquiry and research design, ed. C.N. Poth. 2018, London: London : SAGE Publications.

127. Groenewald, T., *A Phenomenological Research Design Illustrated.* International journal of qualitative methods, 2004. **3**(1): p. 42-55.

128. Little, D., *Positivism / by Daniel Little*

edited by Paul Atkinson, Sara Delamont, Alexandru Cernat, Joseph W. Sakshaug & Richard A. Williams, ed. P. Atkinson, et al. 2020: London : SAGE Publications Ltd.

129. Bryman, A., *Social Research Methods*. 2016: Oxford University Press.

130. Schwartz-Shea, P., *Interpretivism / by Peregrine Schwartz-Shea & Dvora Yanow*

edited by Paul Atkinson, Sara Delamont, Alexandru Cernat, Joseph W. Sakshaug & Richard A. Williams, ed. D. Yanow, et al. 2020: London : SAGE Publications Ltd.

131. Hammond, M., *Research Methods : The Key Concepts*. 2nd ed, ed. J. Wellington. 2020: Milton : Taylor & Francis Group.

132. Goodman, R.B., *Pragmatism / edited by Russell B. Goodman*. 2005, London: London : Routledge.

133. Bacon, M., *Pragmatism: An Introduction*. 2014: Polity Press.
134. Moher, D., et al., *Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement.* PLOS Medicine, 2009. **6**(7): p. e1000097.
135. Centre for Reviews and Dissemination - University of York. *Prospero*. 2018 18 December 2018]; Available from: https://www.crd.york.ac.uk/prospero/.
136. Elsevier. *Embase*. 2023 [cited 2023 31 January 2023]; Available from: https://www.elsevier.com/products/embase.
137. National Library of Medicine. *MEDLINE*. 2023 [cited 2023 31 January 2023]; Available from: https://www.nlm.nih.gov/medline/medline_overview.html.
138. National Center for Biotechnology Information. *PubMed*. 2023 [cited 2023 31 January 2023]; Available from: https://pubmed.ncbi.nlm.nih.gov/.
139. Clarivate Analytics. *Web of Science*. 2023; Available from: https://clarivate.com/webofsciencegroup/solutions/web-of-science/.
140. Gavi. *Publications*. 2023 [cited 2023 31 January 2023]; Available from: https://www.gavi.org/news-resources/publications.
141. OpenGrey. *OPENGREY.EU – Grey Literature Database*. 2023 [cited 2023 31 January 2023]; Available from: http://www.opengrey.eu/.
142. World Health Organization. *WHO Library Catalog*. 2023 [cited 2023 31 January 2023]; Available from: https://kohahq.searo.who.int.
143. Wells G, S.B., O'Connell D, Peterson J, Welch V, Losos M, Tugwell P, . *The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses*. 2018; Available

from:
http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp.

144. Modesti, P.A., et al., *Panethnic Differences in Blood Pressure in Europe: A Systematic Review and Meta-Analysis.* PLoS One, 2016. **11 (Suppl. 1)**(1): p. e0147601.

145. Critical Appraisal Skills Programme. *CASP Checklist*. 2023 [cited 2023 31 January 2023]; Available from: https://casp-uk.net/casp-tools-checklists/.

146. Downs, S.H. and N. Black, *The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions.* Journal of epidemiology and community health (1979), 1998. **52**(6): p. 377-384.

147. Joanna Briggs Institute. *Critical Appraisal Tools*. 2023 [cited 2023 31 January 2023]; Available from: https://jbi.global/critical-appraisal-tools.

148. Sterne, J.A.C., et al., *ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions.* BMJ (Online), 2016. **355**: p. i4919-i4919.

149. Guyatt, G., et al., *GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables.* J Clin Epidemiol, 2011. **64**(4): p. 383-94.

150. Schünemann H, B.J., Guyatt G, Oxman A, editor(s), *Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach (updated October 2013)*. 2013: GRADE Working Group.

151. Berkman, N.D., et al., *AHRQ Methods for Effective Health Care Grading the Strength of a Body of Evidence When Assessing Health Care Interventions for the Effective Health Care Program of the Agency for*

Healthcare Research and Quality: An Update, in *Methods Guide for Effectiveness and Comparative Effectiveness Reviews*. 2008, Agency for Healthcare Research and Quality (US): Rockville (MD).

152. OCEBM Levels of Evidence Working Group. *OCEBM Levels of Evidence*. 2011 [cited 2023 31 January 2023]; Available from: https://www.cebm.ox.ac.uk/resources/levels-of-evidence/ocebm-levels-of-evidence.
153. Higgins, J.P.T., et al., *Cochrane Handbook for Systematic Reviews of Interventions*. Second edition ed. 2019, Newark: Newark: John Wiley & Sons, Incorporated.
154. World Health Organization, *WHO handbook for guideline development*. 2nd ed. 2014, Geneva: World Health Organization,.
155. Davies, H.T., *Interpreting measures of treatment effect.* Hosp Med, 1998. **59**(6): p. 499-501.
156. Efthimiou, O., *Practical guide to the meta-analysis of rare events.* Evid Based Ment Health, 2018. **21**(2): p. 72-76.
157. Brockwell, S.E. and I.R. Gordon, *A comparison of statistical methods for meta-analysis.* Statistics in Medicine, 2001. **20**(6): p. 825-840.
158. Mantel, N. and W. Haenszel, *Statistical aspects of the analysis of data from retrospective studies of disease.* J Natl Cancer Inst, 1959. **22**(4): p. 719-48.
159. Lee, C.H., et al., *Comparison of Two Meta-Analysis Methods: Inverse-Variance-Weighted Average and Weighted Sum of Z-Scores.* Genomics Inform, 2016. **14**(4): p. 173-180.
160. Higgins, J.P.T. and S.G. Thompson, *Quantifying heterogeneity in a meta-analysis.* Statistics in medicine, 2002. **21**(11): p. 1539-1558.
161. Sutton, A.J. and K.R. Abrams, *Bayesian methods in meta-analysis and evidence synthesis.* Statistical methods in medical research, 2001. **10**(4): p. 277-303.

162. Lewis, S. and M. Clarke, *Forest plots: trying to see the wood and the trees.* Bmj, 2001. **322**(7300): p. 1479-80.

163. Higgins, J.P.T., et al., *Measuring inconsistency in meta-analyses.* BMJ, 2003. **327**(7414): p. 557.

164. Allam, R.R., et al., *A Case-control Study of Diphtheria in the High Incidence City of Hyderabad, India.* Pediatr Infect Dis J, 2016. **35**(3): p. 253-6.

165. Bisgard, K.M., et al., *Diphtheria toxoid vaccine effectiveness: A case-control study in Russia.* The Journal of infectious diseases, 2000. **181 Suppl 1**: p. S184-7.

166. Bitragunta, S., et al., *Persistence of diphtheria, Hyderabad, India, 2003-2006.* Emerging Infectious Diseases, 2008. **14**(7): p. 1144-1146.

167. Brennan, M., et al., *How many doses of diphtheria toxoid are required for protection in adults? Results of a case-control study among 40- to 49-year-old adults in the Russian Federation.* Journal of Infectious Diseases, 2000. **181**(SUPPL. 1): p. S193-S196.

168. Chen, R.T., et al., *Ukraine, 1992: First assessment of diphtheria vaccine effectiveness during the recent resurgence of diphtheria in the former Soviet Union.* Journal of Infectious Diseases, 2000. **181**: p. S178-S183.

169. Faria, M.d.A., et al., *[Epidemiological study of diphtheria in greater Sao Paulo, 1969].* Estudo epidemiologico de difteria na regiao de grande Sao Paulo, 1969., 1971. **5**(2): p. 213-20.

170. Husada, D., et al., *Risk Factors of Diphtheria Carriers in Indonesian Children.* Southeast Asian Journal of Tropical Medicine and Public Health, 2018. **49**(4): p. 660-669.

171. Jones, E.E., R.J. Kim-Farley, and M. Algunaid, *Diphtheria: A possible foodborne outbreak in Hodeida, Yemen Arab Republic.* Bulletin of the World Health Organization, 1985. **63**(2): p. 287-293.

172. Murakami, H., et al., *Endemic diphtheria in Ho Chi Minh City; Viet Nam: A matched case-control study to identify risk factors of incidence.* Vaccine, 2010. **28**(51): p. 8141-8146.

173. Nassar, A.A.H., M. Abdullah Al-Amad, and Y.A. Ghaleb, *Risk factors for diphtheria in Sana'a, Yemen, 2019: a matched case-control study.* IJID Reg, 2022. **2**: p. 40-44.

174. Quick, M.L., et al., *Risk factors for diphtheria: A prospective case-control study in the Republic of Georgia, 1995-1996.* Journal of Infectious Diseases, 2000. **181**(SUPPL. 1): p. S121-S129.

175. Ramdan, I.M., et al., *Risk factors for diphtheria outbreak in children aged 1-10 years in East Kalimantan Province, Indonesia.* F1000Res, 2018. **7**: p. 1625.

176. Sein, C., et al., *Diphtheria outbreak in Lao People's Democratic Republic, 2012-2013.* Vaccine, 2016. **34**(36): p. 4321-4326.

177. Vitek, C.R., et al., *Risk of diphtheria among schoolchildren in the Russian Federation in relation to time since last vaccination.* Lancet, 1999. **353**(9150): p. 355-358.

178. Chandra, R., et al., *A follow-up study on diphtheria carriers among children below 5 years in a rural population of Lucknow District.* The Indian journal of medical research, 1973. **61**(3): p. 442-8.

179. Belsey, M.A., et al., *Corynebacterium Diphtheriae Skin Infections in Alabama and Louisiana - a Factor in Epidemiology of Diphtheria.* New England Journal of Medicine, 1969. **280**(3): p. 135-&.

180. Harnisch, J.P., et al., *Diphtheria among alcoholic urban adults. A decade of experience in Seattle.* Annals of internal medicine, 1989. **111**(1): p. 71-82.

181. Kalapothaki, V., T. Sapounas, and E. Xirouchaki, *Prevalence of diphtheria carriers in a population with disappearing clinical diphtheria.* Infection, 1984. **12**(6): p. 387-389.

182. Kitamura, N., et al., *Seroepidemiology and Carriage of Diphtheria in Epidemic-Prone Area and Implications for Vaccination Policy, Vietnam.* Emerg Infect Dis, 2023. **29**(1): p. 70-80.

183. Marcuse, E.K. and M.G. Grand, *Epidemiology of diphtheria in San Antonio, Tex., 1970.* JAMA, 1973. **224**(3): p. 305-10.

184. Miller, L.W.O., J. J.; Drake, J.; et al,, *Diphtheria Immunization. Effect Upon Carriers and the Control of Outbreaks.* American Journal of Diseases of Children, 1972. **123**(3): p. 197-199.

185. Ohuabunwo, C., et al., *Respiratory diphtheria among highly vaccinated military trainees in Latvia: Improved protection from DT compared with Td booster vaccination.* Scandinavian Journal of Infectious Diseases, 2005. **37**(11-12): p. 813-820.

186. Trichopoulos, D., et al., *Diphtheria Carriers among Schoolchildren in Athens.* Scandinavian Journal of Infectious Diseases, 1972. **4**(3): p. 197-201.

187. Coleman, S., *The association between tuberculosis and diphtheria.* Epidemiology and Infection, 2018. **146**(8): p. 940-945.

188. Dureab, F., et al., *Diphtheria outbreak in Yemen: the impact of conflict on a fragile health system.* Confl Health, 2019. **13**: p. 19.

189. Ikejezie, J., et al., *The epidemiology of diphtheria in Haiti, December 2014-June 2021: A spatial modeling analysis.* PLoS One, 2022. **17**(8): p. e0273398.

190. Izza, N. and Soenarnatalina, *Analysis of Spatial Data of Diphtheria Disease in East Java Province during the year 2010 and 2011.* Buletin Penelitian Sistem Kesehatan, 2015. **18**(2): p. 211-219.

191. Podavalenko, A.P., *Estimating the complication risk of epidemic situation with diphtheria in Ukraine.* Asian Journal of Epidemiology, 2018. **11**(1): p. 26-33.

192. Quesada, G.M., et al., *Risk Analysis of the 1970 San-Antonio Diphtheria Epidemic.* Disasters, 1978. **2**(4): p. 221-230.

193. Feery, B.J., et al., *The incidence and type of reactions to plain and adsorbed DTP vaccines.* Australian Paediatric Journal, 1985. **21**(2): p. 91-95.

194. Alexander, P.E., et al., *World Health Organization recommendations are often strong based on low confidence in effect estimates.* Journal of Clinical Epidemiology, 2014. **67**(6): p. 629-634.

195. Fleming, P.S., et al., *High quality of the evidence for medical and other health-related interventions was uncommon in Cochrane systematic reviews.* Journal of Clinical Epidemiology, 2016. **78**: p. 34-42.

196. Faraoni, D. and S.T. Schaefer, *Randomized controlled trials vs. observational studies: why not just live together?* BMC Anesthesiol, 2016. **16**(1): p. 102.

197. Fernainy, P., et al., *Rethinking the pros and cons of randomized controlled trials and observational studies in the era of big data and advanced methods: a panel discussion.* BMC Proceedings, 2024. **18**(2): p. 1.

198. Albhaisi, S. and R.P. Wenzel, *The Value of Medical Registries and Observational Studies Early in Pandemics: The Coronavirus Disease 2019 (COVID-19) Experience.* Clinical Infectious Diseases, 2022. **74**(6): p. 1112-1116.

199. Davies, R., et al., *A systematic review of observational methods used to quantify personal protective behaviours among members of the public during the COVID-19 pandemic, and the concordance between*

observational and self-report measures in infectious disease health protection. BMC Public Health, 2022. **22**(1): p. 1436.

200. Concato, J., N. Shah, and R.I. Horwitz, *Randomized, controlled trials, observational studies, and the hierarchy of research designs.* New England journal of medicine, 2000. **342**(25): p. 1887-1892.

201. Hannan, E.L., *Randomized Clinical Trials and Observational Studies: Guidelines for Assessing Respective Strengths and Limitations.* JACC: Cardiovascular Interventions, 2008. **1**(3): p. 211-217.

202. Abadi, G. and D. Abel Fekadu, *Being unvaccinated and having a contact history increased the risk of measles infection during an outbreak: a finding from measles outbreak investigation in rural district of Ethiopia.* BMC Infectious Diseases, 2019. **19**(1): p. 1-6.

203. Phadke, V.K., et al., *Association Between Vaccine Refusal and Vaccine-Preventable Diseases in the United States: A Review of Measles and Pertussis.* JAMA, 2016. **315**(11): p. 1149-1158.

204. Sitepu, F.Y., et al., *Being unvaccinated and contact with measles cases as the risk factors of measles outbreak, North Sumatera, Indonesia.* Clinical Epidemiology and Global Health, 2019.

205. Vemula, V.N., et al., *Risk factors and clinical profile of measles infection in children in Singapore.* Infection, Disease & Health, 2016. **21**(4): p. 192-196.

206. Mueller, J.K., T; Nwakamma, I; Diallo, C; Yaro, S., *Comparative efficacy/effectiveness of schedules in infant immunisation against pertussis, diphtheria and tetanus: systematic review and meta-analysis*. 2014, École des Hautes Études en Santé Publique;: Rennes, France.

207. Nailul, I. and S. Soenarnatalina, *Analysis of Spatial Data of Diphtheria Disease in East Java Province during the year 2010 and 2011.* Buletin Penelitian Sistem Kesehatan, 2015. **18**(2): p. 211-219.

208. Ang, L.W., L. James, and K.T. Goh, *Prevalence of diphtheria and tetanus antibodies among adults in Singapore: a national serological study to identify most susceptible population groups.* Journal of Public Health, 2015. **38**(1): p. 99-105.

209. Dragomirescu, C.C., et al., *Seroprevalence study of anti diphtheria antibodies in two age-groups of Romanian adults.* Roum Arch Microbiol Immunol, 2014. **73**(1-2): p. 18-24.

210. Wanlapakorn, N., et al., *Diphtheria outbreak in Thailand, 2012; seroprevalence of diphtheria antibodies among Thai adults and its implications for immunization programs.* Southeast Asian J Trop Med Public Health, 2014. **45**(5): p. 1132-41.

211. Zasada, A.A., et al., *Seroprevalence of diphtheria toxoid IgG antibodies in children, adolescents and adults in Poland.* BMC Infectious Diseases, 2013. **13**(1): p. 551.

212. Dimitrios, C.C., et al., *Vaccination Programs for Adults in Europe, 2019.* Vaccines, 2020. **8**(1): p. 34.

213. World Health Organization. *WHO vaccine-preventable diseases: monitoring system. 2020 global summary reference time series: diphtheria*. 2020 [cited 2020; Available from: https://apps.who.int/immunization_monitoring/globalsummary/countries?commit=OK.

214. Abubakar, A., et al., *Outbreak of suspected pertussis in Kaltungo, Gombe State, Northern Nigeria, 2015: the role of sub-optimum routine immunization coverage.* Pan Afr Med J, 2019. **32**(Suppl 1): p. 9.

215. Almaw, L. and H. Bizuneh, *Pertussis outbreak investigation in Janamora district, Amhara Regional State, Ethiopia: a case-control study.* Pan Afr Med J, 2019. **34**: p. 65.

216. Abdulkadir, A. and T.T. Gebrehiwot, *Risk Factors for Rubella Transmission in Kuyu District, Ethiopia, 2018: A Case-Control Study.* Interdisciplinary Perspectives on Infectious Diseases, 2019. **2019**.

217. Huang, A.S., et al., *Risk Factors for Mumps at a University with a Large Mumps Outbreak.* Public Health Reports, 2009. **124**(3): p. 419-426.

218. Yaqub, O., et al., *Attitudes to vaccination: A critical review.* Social Science & Medicine, 2014. **112**: p. 1-11.

219. Brieger, D., et al., *Knowledge, attitudes and opinions towards measles and the MMR vaccine across two NSW cohorts.* Australian and New Zealand Journal of Public Health, 2017. **41**(6): p. 641-646.

220. Qutaiba B Al-Lela, O., et al., *Are parents' knowledge and practice regarding immunization related to pediatrics' immunization compliance? a mixed method study.* BMC pediatrics, 2014. **14**(1): p. 20.

221. Larson, H.J., et al., *Addressing the vaccine confidence gap.* The Lancet, 2011. **378**(9790): p. 526-535.

222. Dube, E., et al., *Vaccine hesitancy: an overview.* Hum Vaccin Immunother, 2013. **9**(8): p. 1763-73.

223. Rabie, T. and V. Curtis, *Handwashing and risk of respiratory infections: a quantitative systematic review.* Trop Med Int Health, 2006. **11**(3): p. 258-67.

224. Aiello, A.E., et al., *Effect of hand hygiene on infectious disease risk in the community setting: a meta-analysis.* American journal of public health, 2008. **98**(8): p. 1372-1381.

225. Larson, H.J., et al., *Understanding vaccine hesitancy around vaccines and vaccination from a global perspective: A systematic review of published literature, 2007–2012.* Vaccine, 2014. **32**(19): p. 2150-2159.

226. Rainey, J.J., et al., *Reasons related to non-vaccination and under-vaccination of children in low and middle income countries: findings from a systematic review of the published literature, 1999-2009.* Vaccine, 2011. **29**(46): p. 8215-21.

227. Vikram, K., R. Vanneman, and S. Desai, *Linkages between maternal education and childhood immunization in India.* Social Science & Medicine, 2012. **75**(2): p. 331-339.

228. LeVine, R.A. and M.L. Rowe, *Maternal Literacy and Child Health in Less-Developed Countries: Evidence, Processes, and Limitations.* 2009. **30**(4).

229. Glewwe, P., *Why does mother's schooling raise child health in developing countries? Evidence from Morocco.* Journal Of Human Resources, 1999. **34**(1): p. 124-159.

230. Joshi, A.R., *Maternal schooling and child health: preliminary analysis of the intervening mechanisms in rural Nepal.* Health transition review : the cultural, social, and behavioural determinants of health, 1994. **4**(1): p. 1.

231. Desai, S. and S. Alva, *Maternal education and child health: Is there a strong causal relationship?* Demography, 1998. **35**(1): p. 71-81.

232. Frost, M.B., R. Forste, and D.W. Haas, *Maternal education and child nutritional status in Bolivia: finding the links.* Soc Sci Med, 2005. **60**(2): p. 395-407.

233. Blewett, L.A., et al., *The impact of gaps in health insurance coverage on immunization status for young children.* Health Serv Res, 2008. **43**(5 Pt 1): p. 1619-36.

234. Favin, M., et al., *Why children are not vaccinated: a review of the grey literature.* Int Health, 2012. **4**(4): p. 229-38.

235. Vitek, C.R., *Diphtheria*, in *Mass Vaccination: Global Aspects — Progress and Obstacles*, S.A. Plotkin, Editor. 2006, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 71-94.

236. Patel, S. and R. Dowse, *Understanding the medicines information-seeking behaviour and information needs of South African long-term patients with limited literacy skills.* Health Expectations, 2015. **18**(5): p. 1494-1507.

237. Kuske, S., et al., *Diabetes-related information-seeking behaviour: a systematic review.* Systematic reviews, 2017. **6**(1): p. 212-212.

238. Tang, C., et al., *Examining income-related inequality in health literacy and health-information seeking among urban population in China.* BMC Public Health, 2019. **19**(1): p. 221.

239. Agyemang-Duah, W., et al., *Dynamics of health information-seeking behaviour among older adults with very low incomes in Ghana: a qualitative study.* BMC Public Health, 2020. **20**(1): p. 928.

240. Obregon, R., et al., *Achieving polio eradication: a review of health communication evidence and lessons learned in India and Pakistan.* Bull World Health Organ, 2009. **87**(8): p. 624-30.

241. Obregón, R. and S. Waisbord, *The Complexity of Social Mobilization in Health Communication: Top-Down and Bottom-Up Experiences in Polio Eradication.* Journal of Health Communication, 2010. **15**(sup1): p. 25-47.

242. Waisbord, S., et al., *Communication for Polio Eradication: Improving the Quality of Communication Programming Through Real-Time Monitoring and Evaluation.* Journal of Health Communication, 2010. **15**(sup1): p. 9-24.

243. McArthur-Lloyd, A., et al., *Community Engagement, Routine Immunization, and the Polio Legacy in Northern Nigeria.* Global Health Communication, 2016. **2**(1): p. 1-10.

244. Franzen, S.R., C. Chandler, and T. Lang, *Health research capacity development in low and middle income countries: reality or rhetoric? A systematic meta-narrative review of the qualitative literature.* BMJ Open, 2017. **7**(1): p. e012332.

245. World Health Organization, *Research for universal health coverage*. 2013: Geneva, Switzerland : World Health Organization.

246. Harle, J.W., V, *Open Access: challenges and opportunities for Low- and Middle-Income Countries and the potential impact of UK policy*. 2020, London: Foreign, Commonwealth & Development Office.

247. Smith, A.C., et al., *Assessing the effect of article processing charges on the geographic diversity of authors using Elsevier's "Mirror Journal" system.* Quantitative Science Studies, 2021. **2**(4): p. 1123-1143.

248. Røttingen, J.-A.P., et al., *Mapping of available health research and development data: what's there, what's missing, and what role is there for a global observatory?* The Lancet (British edition), 2013. **382**(9900): p. 1286-1307.

249. Causey, K., et al., *Estimating global and regional disruptions to routine childhood vaccine coverage during the COVID-19 pandemic in 2020: a modelling study.* Lancet, 2021. **398**(10299): p. 522-534.

250. Shet, A., et al., *Impact of the SARS-CoV-2 pandemic on routine immunisation services: evidence of disruption and recovery from 170 countries and territories.* Lancet Glob Health, 2022. **10**(2): p. e186-e194.

251. Althubaiti, A., *Information bias in health research: definition, pitfalls, and adjustment methods.* J Multidiscip Healthc, 2016. **9**: p. 211-7.

252. Krumpal, I., *Determinants of social desirability bias in sensitive surveys: a literature review.* Quality & quantity, 2013. **47**(4): p. 2025-2047.

253. Gianfrancesco, M.A. and N.D. Goldstein, *A narrative review on the validity of electronic health record-based research in epidemiology*. BMC Medical Research Methodology, 2021. **21**(1): p. 234.

254. Hersh, W.R., et al., *Caveats for the use of operational electronic health record data in comparative effectiveness research.* Med Care, 2013. **51**(8 Suppl 3): p. S30-7.

255. Clerville, J., *Diphtheria Outbreak, Haiti, 2014-2017: An epidemiological profile and a case fatality rate trend analysis*, in *18th International Congress on Infectious Diseases*. 2018: Buenos Aires

256. Quesada, G.M., et al., *Risk analysis of the 1970 San Antonio diphtheria epidemic.* Disasters, 1978. **2**(4): p. 221-230.

257. Haitian Institute of Statistics and Informatics, *Total population, population aged 18 years and over, estimated households and densities in 2015*, Directorate of Demographic and Social Statistics, Editor. 2015, Haitian Institute of Statisitics and Informatics,: Port-au-Prince.

258. Graves, K.A., *Haiti*. 2006: Capstone.

259. United States Central Intelligence Agency, *The 2017 CIA World Factbook*. 2017: Project Gutenberg.

260. UN OCHA, *Hurricane Matthew situation report no. 20.* 2016, UN OCHA,: Port-au-Prince, Haiti.

261. United States Central Intelligence Agency. *Haiti*. 2022 [cited 2023 6 January 2023]; Available from: https://www.cia.gov/the-world-factbook/countries/haiti/.

262. Lansford, T., *Political Handbook of the World 2020-2021*. 2021: SAGE Publications.

263. Mèrancourt, W. and A. Coletta, *Senators' departure leaves Haiti without an elected government*, in *The Washington Post*. 2023. p. NA.

264. United Nations Development Programme. *Human Development Index (HDI)*. 2023 [cited 2023 6 January 2023]; Available from: https://hdr.undp.org/data-center/human-development-index#/indicies/HDI.

265. The World Bank. *GDP per capita (current US$) - Latin America & Caribbean*. 2023 [cited 2023; Available from: https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?locations=ZJ.

266. The World Bank, *Haiti - COVID-19 Strategic Preparedness and Response Project* 2020: Washington, D.C.

267. The World Bank. *Unemployment, total (% of total labor force) (modeled ILO estimate) - Haiti*. 2023 [cited 2023; Available from: https://data.worldbank.org/indicator/SL.UEM.TOTL.ZS?locations=HT.

268. Bhattacharya, R. and N. Shenai, *Lessons from Haiti's Recent Exchange Rate Developments. IMF Working Paper No. 2022/225.* 2022.

269. Juin, S., et al., *Strengthening National Disease Surveillance and Response-Haiti, 2010-2015.* The American journal of tropical medicine and hygiene, 2017. **97**(4_Suppl): p. 12-20.

270. Juin, S., et al., *Strengthening National Disease Surveillance and Response-Haiti, 2010-2015.* Am J Trop Med Hyg, 2017. **97**(4_Suppl): p. 12-20.

271. The World Bank, *Better Spending, Better Care*. Health Sector Review. 2017: The World Bank,. -1.

272. World Health Organization, U. *Immunization dashboard*. 2022 6 January 2023]; Available from: https://immunizationdata.who.int.

273. Magloire, R., et al., *Launching a National Surveillance System After an Earthquake—Haiti, 2010.* JAMA : the journal of the American Medical Association, 2010. **304**(12): p. 1318-1320.

274. Ikejezie, J., et al., *Modifiable risk factors for diphtheria: a systematic review and meta-analysis.* Manuscript in preparation, 2022.

275. ICF. *The DHS Program Spatial Data Repository. Funded by USAID.* 2019 [cited 2019; Available from: spatialdata.dhsprogram.com.

276. Haitian Institute of Childhood (IHE), H.I.o.S.a.I., ICF International, Ministry of Public Health and Population (Haiti),, *Haiti Demographic and Health Survey 2016-2017*. 2018, Fairfax, United States of America: ICF International.

277. R Development Core Team, *R: A language and environment for statistical computing*. 2023, Vienna: R Foundation for Statistical Computing,.

278. Humanitarian Data Exchange. *Haiti*. 2019 [cited 2021; Available from: https://data.humdata.org/group/hti.

279. QGIS Development Team. *QGIS Geographic Information System*. 2019 [cited 2019; Available from: http://qgis.osgeo.org.

280. Moran, P.A.P., *A Test for the Serial Independence of Residuals.* Biometrika, 1950. **37**(1/2): p. 178-181.

281. Mitchell, A., *The ESRI guide to GIS analysis, Volume 2*. Spatial Measurements & Statistics, ed. I. Environmental Systems Research. 2005, Redlands, Calif.: Redlands, Calif. : Environmental Systems Research Institute.

282. Anselin, L., *Local Indicators of Spatial Association—LISA.* Geographical Analysis, 1995. **27**(2): p. 93-115.

283. Ward, M.D.G., Kristian Skrede *Spatial regression models*, ed. K.S. Gleditsch. 2008, Thousand Oaks: Thousand Oaks : Sage Publications.

284. Ripley, B.D., *The second-order analysis of stationary point processes.* Journal of applied probability, 1976. **13**(2): p. 255-266.

285. Kulldorff, M., *A spatial scan statistic.* Communications in statistics. Theory and methods, 1997. **26**(6): p. 1481-1496.

286. Getis, A. and J.K. Ord, *The Analysis of Spatial Association by Use of Distance Statistics.* Geographical analysis, 1992. **24**(3): p. 189-206.

287. Fotheringham, A.S. and P. Rogerson, *The SAGE handbook of spatial analysis [electronic resource] / edited by A. Stewart Fotheringham and Peter A. Rogerson*. Handbook of spatial analysis. 2009, Los Angeles, [Calif.]

London: Los Angeles, Calif.

London : SAGE.

288. Anselin, L., I. Syabri, and Y. Kho, *GeoDa : An Introduction to Spatial Data Analysis.* Geographical Analysis, 2006. **38**(1): p. 5-22.

289. Ripley, B.D., *Spatial statistics*. Vol. 575. 2005: John Wiley & Sons.

290. Stewart Fotheringham, A., M. Charlton, and C. Brunsdon, *The geography of parameter space: an investigation of spatial non-stationarity.* International Journal of Geographical Information Systems, 1996. **10**(5): p. 605-627.

291. Brunsdon, C., A.S. Fotheringham, and M.E. Charlton, *Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity.* Geographical Analysis, 1996. **28**(4): p. 281-298.

292. Tobler, W.R., *A Computer Movie Simulating Urban Growth in the Detroit Region.* Economic Geography, 1970. **46**: p. 234-240.

293. *The problem of multicollinearity*, in *Understanding Regression Analysis*, M.P. Allen, Editor. 1997, Springer US: Boston, MA. p. 176-180.

294. Forthofer, R.N., *Biostatistics : a guide to design, analysis, and discovery / Ronald N. Forthofer, Eun Sul Lee, Mike Hernandez*. 2nd

ed. ed. Introduction to biostatistics : a guide to design, analysis, and discovery, ed. E.S. Lee and M. Hernandez. 2007, Burlington, Mass.

London: Burlington, Mass.

London : Elsevier Academic Press.

295. Akaike, H., *Information Theory and an Extension of the Maximum Likelihood Principle.* Breakthroughs in Statistics, 2015: p. 610-624.

296. LeSage, J. and R.K. Pace, *Introduction to Spatial Econometrics*. Statistics: A Series of Textbooks and Monographs. 2009: CRC Press.

297. Anselin, L., *Spatial Econometrics: Methods and Models*. Studies in Operational Regional Science. 2013: Springer Netherlands.

298. Banerjee, S., B.P. Carlin, and A.E. Gelfand, *Hierarchical Modeling and Analysis for Spatial Data, Second Edition*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. 2014: Taylor & Francis.

299. Gelman, A., et al., *Bayesian Data Analysis, Third Edition*. Chapman & Hall/CRC Texts in Statistical Science. 2013: Taylor & Francis.

300. Wheeler, D.C., *Geographically Weighted Regression*, in *Handbook of Regional Science*, M.M. Fischer and P. Nijkamp, Editors. 2014, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 1435-1459.

301. Minta, A.A., et al., *Seroprevalence of Measles, Rubella, Tetanus, and Diphtheria Antibodies among Children in Haiti, 2017.* Am J Trop Med Hyg, 2020. **103**(4): p. 1717-1725.

302. Pan American Health Organization / World Health Organization. *Haiti launches campaign to vaccinate over 2 million children against diphtheria, with PAHO support*. 2018; Available from: https://www.paho.org/en/news/10-4-2018-haiti-launches-

campaign-vaccinate-over-2-million-children-against-diphtheria-paho.

303. United Nations Integrated Office in Haiti, *Report of the Secretary-General - S/2020/123*. 2020: Port-au-Prince, Haiti.

304. World Health Organization, *Coronavirus Disease 2019 (COVID-19) Situation Report–61*. 2020: Geneva, Switzerland.

305. Pan American Health Organization / World Health Organization, *Epidemiological Update: Diphtheria in Hispaniola. 25 June 2021*. 2021: Washington, D.C.

306. Holm, M. and H. Burkhartzmeyer, *Implementation of a phased medical educational approach in a developing country.* Glob Health Action, 2015. **8**: p. 29882.

307. Louis, F.J., et al., *Specimen Referral Network to Rapidly Scale-Up CD4 Testing: The Hub and Spoke Model for Haiti.* J AIDS Clin Res, 2015. **6**(8).

308. Séraphin, M.N., et al., *Childhood anemia in Rural Haiti: the potential role of community health workers.* Global Health Research and Policy, 2017. **2**(1): p. 3.

309. Nelli, L., et al., *Distance sampling for epidemiology: an interactive tool for estimating under-reporting of cases from clinic data.* International Journal of Health Geographics, 2020. **19**(1): p. 16.

310. Hierink, F., et al., *The winding road to health: A systematic scoping review on the effect of geographical accessibility to health care on infectious diseases in low- and middle-income countries.* PLOS ONE, 2021. **16**(1): p. e0244921.

311. World Health Organization, *Assessing tuberculosis under-reporting through inventory studies*. 2012, Geneva: World Health Organization,.

312. Dye, C., et al., *Measuring tuberculosis burden, trends, and the impact of control programmes.* The Lancet Infectious Diseases, 2008. **8**(4): p. 233-243.

313. MacPherson, P., et al., *Community-based active-case finding for tuberculosis: navigating a complex minefield.* BMC Global and Public Health, 2024. **2**(1): p. 9.

314. Onozaki, I., et al., *National tuberculosis prevalence surveys in Asia, 1990–2012: an overview of results and lessons learned.* Tropical Medicine & International Health, 2015. **20**(9): p. 1128-1145.

315. Saker, L., et al., *Globalization and infectious diseases : a review of the linkages / Lance Saker ... [et al.]*. 2004, World Health Organization: Geneva.

316. Norwegian Institute of Public Health, *Urbanization and preparedness for outbreaks with high-impact respiratory pathogens*. 2020: Oslo, Norway.

317. Braveman, P. and L. Gottlieb, *The social determinants of health: it's time to consider the causes of the causes.* Public health reports (Washington, D.C. : 1974), 2014. **129 Suppl 2**(Suppl 2): p. 19-31.

318. Marmot, M. and R. Wilkinson, *Social Determinants of Health*. 2005, Oxford: Oxford: Oxford University Press.

319. Chekol, Y.M., et al., *Geographic weighted regression analysis of hot spots of modern contraceptive utilization and its associated factors in Ethiopia.* PLOS ONE, 2023. **18**(11): p. e0288710.

320. Stashko, L.A., et al., *Assessing the quality and accuracy of national immunization program reported target population estimates from 2000 to 2016.* PLOS ONE, 2019. **14**(7): p. e0216933.

321. Bosch-Capblanch, X., et al., *Accuracy and quality of immunization information systems in forty-one low income countries.* Tropical Medicine & International Health, 2009. **14**(1): p. 2-10.

322. The World Bank, *International development association project appraisal document on a proposed grant in the amount of sdr 3.8 million (US$5 million equivalent) to the Republic of Haiti for a statistical capacity building project*. 2017, The World Bank.

323. Robinson, W.S., *Ecological Correlations and the Behavior of Individuals.* International journal of epidemiology, 2009. **38**(2): p. 337-341.

324. Shi, H., L. Zhang, and J. Liu, *A new spatial-attribute weighting function for geographically weighted regression.* Canadian journal of forest research, 2006. **36**(4): p. 996-1005.

325. Wheeler, D.C., *Diagnostic Tools and a Remedial Method for Collinearity in Geographically Weighted Regression.* Environment and planning. A, 2007. **39**(10): p. 2464-2481.

326. Eysenbach, G., *Infodemiology and infoveillance tracking online health information and cyberbehavior for public health.* Am J Prev Med, 2011. **40**(5 Suppl 2): p. S154-8.

327. Statista. *Twitter: number of monetizable daily active users worldwide 2017-2022*. 2022 [cited 2022 30 May 2022]; Available from: https://www.statista.com/statistics/970920/monetizable-daily-active-twitter-users-worldwide/.

328. Murthy, D., *Twitter : social communication in the Twitter age / Dhiraj Murthy*. 2013, Cambridge: Cambridge : Polity.

329. Signorini, A., A.M. Segre, and P.M. Polgreen, *The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic.* PloS one, 2011. **6**(5): p. e19467-e19467.

330. Househ, M., *Communicating Ebola through social media and electronic news media outlets: A cross-sectional study.* Health informatics journal, 2016. **22**(3): p. 470-478.

331. Shin, S.-Y., et al., *High correlation of Middle East respiratory syndrome spread with Google search and Twitter trends in Korea.* Scientific reports, 2016. **6**(1): p. 32920-32920.

332. Pruss, D., et al., *Zika discourse in the Americas: A multilingual topic analysis of Twitter.* PloS one, 2019. **14**(5): p. e0216922.

333. Karmegam, D. and B. Mapillairaju, *What people share about the COVID-19 outbreak on Twitter? An exploratory analysis.* BMJ health & care informatics, 2020. **27**(3).

334. Dyar, O.J., E. Castro-Sánchez, and A.H. Holmes, *What makes people talk about antibiotics on social media? A retrospective analysis of Twitter use.* J Antimicrob Chemother, 2014. **69**(9): p. 2568-72.

335. Landau, S., et al., *Cluster Analysis*. 2011, United Kingdom: John Wiley & Sons, Incorporated: United Kingdom.

336. Charmaz, K., *Constructing grounded theory : a practical guide through qualitative analysis / Kathy Charmaz*. 2006, London.: London. : Sage Publications.

337. Team, R.C., *R: A language and environment for statistical computing*. 2023, R Foundation for Statistical Computing.

338. Team, R., *RStudio: Integrated Development Environment for R*. 2023, RStudio, PBC.

339. Silge, J. and D. Robinson, *Text mining with R : a tidy approach*. 2017.

340. World Health Organization. *WHO regional offices*. 2023 [cited 2023; Available from: https://www.who.int/about/who-we-are/regional-offices.

341. The World Bank. *World Bank country and lending groups.* 2023 [cited 2023 28 July 2023]; Available from: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups.

342. Dean, R.T. and W.T. Dunsmuir, *Dangers and uses of cross-correlation in analyzing time series in perception, performance, movement, and neuroscience: The importance of constructing transfer function autoregressive models.* Behav Res Methods, 2016. **48**(2): p. 783-802.

343. Chatfield, C. and H. Xing, *The Analysis of Time Series: An Introduction with R*. 2019: CRC Press.

344. Brian, S.E., et al., *Cluster Analysis, 5th Edition*. 2011: Wiley.

345. Glaser, B.G., *The discovery of grounded theory : strategies for qualitative research / Barney G. Glaser and Anselm L. Strauss*, ed. A.L. Strauss. 1967, New York: New York : Aldine.

346. DePoy, E., *Introduction to research : understanding and applying multiple strategies / Elizabeth DePoy, Laura N. Gitlin*. Sixth edition. ed, ed. L.N. Gitlin. 2020: St. Louis : Mosby.

347. Titscher, S., et al., *Methods of Text and Discourse Analysis: In Search of Meaning*. 1 ed. 2000, London: London: SAGE Publications, Limited.

348. Levy, D., *Qualitative Methodology and Grounded Theory in Property Research.* Pacific rim property research journal, 2006. **12**(4): p. 369-388.

349. Glaser, B.G., *Theoretical sensitivity : advances in the methodology of grounded theory / Barney G. Glaser*. 1978, Mill Valley, Calif.: Mill Valley, Calif. : Sociology Press.

350. Microsoft Corporation, *Microsoft Excel.* 2023, Microsoft Corporation,: Redmond (WA).

351. Berger, R., *Now I see it, now I don't: researcher's position and reflexivity in qualitative research.* Qualitative research : QR, 2015. **15**(2): p. 219-234.

352. Finlay, L., *Negotiating the swamp: the opportunity and challenge of reflexivity in research practice.* Qualitative research : QR, 2002. **2**(2): p. 209-230.

353. Gavin, B.S., *Reflexivity and Subjectivity in Qualitative Research: The Utility of a Wittgensteinian Framework.* Forum, qualitative social research, 2002. **3**(3).

354. Maxwell, J.A., *Qualitative research design : an interactive approach / Joseph A. Maxwell*. 3rd ed. ed. 2013, Los Angeles

London: Los Angeles

London : SAGE Publications.

355. Ridgeway, C.L. and S.J. Correll, *Unpacking the Gender System: A Theoretical Perspective on Gender Beliefs and Social Relations.* Gender & society, 2004. **18**(4): p. 510-531.

356. Denzin, N.K. and Y.S. Lincoln, *The handbook of qualitative research / Norman K. Denzin, Yvonna S. Lincoln, editors*. 2nd ed. ed. 2000, Thousand Oaks, Calif.

London: Thousand Oaks, Calif.

London : Sage.

357. Temple, B. and A. Young, *Qualitative Research and Translation Dilemmas.* Qualitative research : QR, 2004. **4**(2): p. 161-178.

358. Noble, H. and J. Smith, *Issues of validity and reliability in qualitative research.* Evidence-based nursing, 2015. **18**(2): p. 34-35.

359. The World Bank. *Mobile cellular subscriptions (per 100 people)*. 2024 [cited 2024 6 December 2024]; Available from: https://data.worldbank.org/indicator/IT.CEL.SETS.P2.

360. Meadows, C.Z., L. Tang, and W. Liu, *Twitter message types, health beliefs, and vaccine attitudes during the 2015 measles outbreak in California.* American Journal of Infection Control, 2019. **47**(11): p. 1314-1318.

361. Househ, M., *Communicating Ebola through social media and electronic news media outlets: A cross-sectional study.* Health Informatics Journal, 2015. **22**(3): p. 470-478.

362. Kwak, H., et al. *What is Twitter, a social network or a news media?* in *International World Wide Web Conference*. ACM.

363. Zhu, L. and K. Lerman, *Attention inequality in social media.* arXiv preprint arXiv:1601.07200, 2016.

364. Durazzi, F., et al., *Clusters of science and health related Twitter users become more isolated during the COVID-19 pandemic.* Scientific Reports, 2021. **11**(1): p. 19655.

365. Ola, O. and K. Sedig, *Understanding Discussions of Health Issues on Twitter: A Visual Analytic Study.* Online J Public Health Inform, 2020. **12**(1): p. e2.

366. Porat, T., et al., *Content and source analysis of popular tweets following a recent case of diphtheria in Spain.* Eur J Public Health, 2019. **29**(1): p. 117-122.

367. Amores, J.J., D. Blanco-Herrero, and C. Arcila-Calderón *The Conversation around COVID-19 on Twitter—Sentiment Analysis and Topic Modelling to Analyse Tweets Published in English during the First Wave of the Pandemic*. Journalism and Media, 2023. **4**, 467-484 DOI: 10.3390/journalmedia4020030.

368. Pascual-Ferrá, P., N. Alperstein, and D.J. Barnett, *Social Network Analysis of COVID-19 Public Discourse on Twitter: Implications for Risk Communication.* Disaster Medicine and Public Health Preparedness, 2022. **16**(2): p. 561-569.

369. Sleigh, J., et al., *Qualitative analysis of visual risk communication on twitter during the Covid-19 pandemic.* BMC Public Health, 2021. **21**(1): p. 810.

370. Laison, E.K.E., et al., *Identifying Potential Lyme Disease Cases Using Self-Reported Worldwide Tweets: Deep Learning Modeling Approach Enhanced With Sentimental Words Through Emojis.* J Med Internet Res, 2023. **25**: p. e47014.

371. Hagg, E., V.S. Dahinten, and L.M. Currie, *The emerging use of social media for health-related purposes in low and middle-income countries: A scoping review.* Int J Med Inform, 2018. **115**: p. 92-105.

372. Fazel, S., et al., *Harnessing Twitter data to survey public attention and attitudes towards COVID-19 vaccines in the UK.* Scientific Reports, 2021. **11**(1): p. 23402.

373. Yousefinaghani, S., et al., *Credibility of vaccine-related content on Twitter during COVID-19 pandemic.* PLOS Global Public Health, 2023. **3**(7): p. e0001385.

374. Pew Research Center. *Demographics of Mobile Device Ownership and Adoption in the United States.* 2021 [cited 2023; Available from: https://www.pewresearch.org/internet/fact-sheet/mobile/.

375. Charles-Smith, L.E., et al., *Using Social Media for Actionable Disease Surveillance and Outbreak Management: A Systematic Literature Review.* PLOS ONE, 2015. **10**(10): p. e0139701.

376. Cuomo, R.E., et al., *A longitudinal and geospatial analysis of COVID-19 tweets during the early outbreak period in the United States.* BMC Public Health, 2021. **21**(1): p. 793.

377. Owuor, I. and H.H. Hochmair *Temporal Relationship between Daily Reports of COVID-19 Infections and Related GDELT and Tweet Mentions*. Geographies, 2023. **3**, 584-609 DOI: 10.3390/geographies3030031.

378. Yousefinaghani, S., et al., *The Assessment of Twitter's Potential for Outbreak Detection: Avian Influenza Case Study.* Scientific Reports, 2019. **9**(1): p. 18147.

379. Bogdanowicz, A. and C. Guan, *Dynamic topic modeling of twitter data during the COVID-19 pandemic.* PLOS ONE, 2022. **17**(5): p. e0268669.

380. de Melo, T. and C.M.S. Figueiredo, *Comparing News Articles and Tweets About COVID-19 in Brazil: Sentiment Analysis and Topic Modeling Approach.* JMIR Public Health Surveill, 2021. **7**(2): p. e24585.

381. Krishnan, N., et al., *Research note: Examining how various social media platforms have responded to COVID-19 misinformation.* Harvard Kennedy School Misinformation Review, 2021. **2**(6): p. 1-25.

382. Lyu, J.C. and G.K. Luli, *Understanding the Public Discussion About the Centers for Disease Control and Prevention During the COVID-19 Pandemic Using Twitter Data: Text Mining Analysis Study.* J Med Internet Res, 2021. **23**(2): p. e25108.

383. McMann, T., et al., *Detection and Characterization of Web-Based Pediatric COVID-19 Vaccine Discussions and Racial and Ethnic Minority Topics: Retrospective Analysis of Twitter Data.* JMIR Pediatr Parent, 2023. **6**: p. e48004.

384. Verma, M., et al., *People's perceptions on COVID-19 vaccination: an analysis of twitter discourse from four countries.* Scientific Reports, 2023. **13**(1): p. 14281.

385. Pulido, C.M., et al., *COVID-19 infodemic: More retweets for science-based information on coronavirus than for false information.* International Sociology, 2020. **35**(4): p. 377-392.

386. Zhai, S., Y.J. Li, and M. Chi, *The Impact of Government Social Media Information Quality on Public Panic During the Infodemic.* Front Psychol, 2022. **13**: p. 908213.

387. Bonnevie, E., et al., *Using social media influencers to increase knowledge and positive attitudes toward the flu vaccine.* PLoS One, 2020. **15**(10): p. e0240828.

388. Vosoughi, S., D. Roy, and S. Aral, *The spread of true and false news online.* Science, 2018. **359**(6380): p. 1146-1151.

389. X Corp, *Giving you more characters to express yourself*, in *X Blog*, X Corp, Editor. 2017, X Corp,.

390. DePaula, N., et al., *Platform Effects on Public Health Communication: A Comparative and National Study of Message Design and Audience Engagement Across Twitter and Facebook.* JMIR Infodemiology, 2022. **2**(2): p. e40198.

391. Phua, J., S.V. Jin, and J. Kim, *Uses and gratifications of social networking sites for bridging and bonding social capital: A comparison of Facebook, Twitter, Instagram, and Snapchat.* Computers in Human Behavior, 2017. **72**: p. 115-122.

392. Blank, G. and C. Lutz, *Representativeness of Social Media in Great Britain: Investigating Facebook, LinkedIn, Twitter, Pinterest, Google+, and Instagram.* American Behavioral Scientist, 2017. **61**(7): p. 741-756.

393. Greenwood, S., A. Perrin, and M. Duggan, *Social media update 2016.* Pew Research Center, 2016. **11**(2): p. 1-18.

394. Benoit, S.L. and R.F. Mauldin, *The "anti-vax" movement: a quantitative report on vaccine beliefs and knowledge across social media.* BMC Public Health, 2021. **21**(1): p. 2106.

395. Hwang, J., et al., *Vaccine discourse during the onset of the COVID-19 pandemic: Topical structure and source patterns informing efforts to combat vaccine hesitancy.* PLoS One, 2022. **17**(7): p. e0271394.

396. Allen, J., D.J. Watts, and D.G. Rand, *Quantifying the impact of misinformation and vaccine-skeptical content on Facebook.* Science. **384**(6699): p. eadk3451.

397. Loomba, S., et al., *Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA.* Nature Human Behaviour, 2021. **5**(3): p. 337-348.

398. Dittmann, S., et al., *Successful Control of Epidemic Diphtheria in the States of the Former Union of Soviet Socialist Republics: Lessons Learned.* The Journal of Infectious Diseases, 2000. **181**(Supplement_1): p. S10-S22.

399. Zaffran, M., et al., *The imperative for stronger vaccine supply and logistics systems.* Vaccine, 2013. **31**: p. B73-B80.

400. Dudley, L. and P. Garner, *Strategies for integrating primary health services in low- and middle-income countries at the point of delivery.* Cochrane Database of Systematic Reviews, 2011(7).

401. Wirtz, C., et al., *Integrating HPV vaccination programs with enhanced cervical cancer screening and treatment, a systematic review.* Vaccine, 2022. **40 Suppl 1**: p. A116-a123.

402. Gizaw, Z., T. Astale, and G.M. Kassie, *What improves access to primary healthcare services in rural communities? A systematic review.* BMC Primary Care, 2022. **23**(1): p. 313.

403. Olmsted, S.S., et al., *Strengthening Laboratory Systems in Resource-Limited Settings.* American Journal of Clinical Pathology, 2010. **134**(3): p. 374-380.

404. Adedeji, O.J., et al., *Towards Universal Health Coverage in Africa: Relevance of Telemedicine and Mobile Clinics.* South Eastern European journal of public health, 2023.

405. McGowan, C.R., et al., *Mobile clinics in humanitarian emergencies: a systematic review.* Conflict and Health, 2020. **14**(1): p. 4.

406. Alghamdi, N.S. and S.M. Alghamdi, *The Role of Digital Technology in Curbing COVID-19.* Int J Environ Res Public Health, 2022. **19**(14).

407. World Health Organization, *Consolidated telemedicine implementation guide*. 2022, Geneva: World Health Organization,.

408. Fanelli, S., et al., *Insights for the future of health system partnerships in low- and middle-income countries: a systematic literature review.* BMC Health Services Research, 2020. **20**(1): p. 571.

409. Joudyian, N., et al., *Public-private partnerships in primary health care: a scoping review.* BMC Health Services Research, 2021. **21**(1): p. 4.

410. Abiiro, G.A. and M. De Allegri, *Universal health coverage from multiple perspectives: a synthesis of conceptual literature and global debates.* BMC International Health and Human Rights, 2015. **15**(1): p. 17.

411. World Health Organization, *Tracking Universal Health Coverage : 2023 Global Monitoring Report*. 1st ed. 2023: Geneva : World Health Organization.

412. Al-Worafi, Y.M., *Medicines Supply in Developing Countries*, in *Handbook of Medical and Health Sciences in Developing Countries : Education, Practice, and Research*, Y.M. Al-Worafi, Editor. 2023, Springer International Publishing: Cham. p. 1-23.

413. Roth, L., et al., *Expanding global access to essential medicines: investment priorities for sustainably strengthening medical product regulatory systems.* Globalization and Health, 2018. **14**(1): p. 102.

414. De Buck, E., et al., *Promoting handwashing and sanitation behaviour change in low-and middle-income countries: a mixed-method systematic review. 3ie Systematic Review 36.* 2017.

415. Igaki, S., et al., *Effectiveness of community and school-based sanitation interventions in improving latrine coverage: a systematic review and meta-analysis of randomized controlled interventions.* Environmental Health and Preventive Medicine, 2021. **26**(1): p. 26.

416. Ludwick, T., et al., *The distinctive roles of urban community health workers in low- and middle-income countries: a scoping review of the literature.* Health Policy and Planning, 2020. **35**(8): p. 1039-1052.

417. Miller, N.P., et al., *Community health workers in humanitarian settings: Scoping review.* J Glob Health, 2020. **10**(2): p. 020602.

418. Edo-Osagie, O., et al., *A scoping review of the use of Twitter for public health research.* Comput Biol Med, 2020. **122**: p. 103770.

419. Tseole, N.P., et al., *Barriers and facilitators to Water, Sanitation and Hygiene (WaSH) practices in Southern Africa: A scoping review.* PLOS ONE, 2022. **17**(8): p. e0271726.

420. World Health, O., *Guidelines for drinking-water quality: small water supplies*. 2024, Geneva: World Health Organization.

421. Sharma Waddington, H., et al., *Impact on childhood mortality of interventions to improve drinking water, sanitation, and hygiene (WASH) to households: Systematic review and meta-analysis.* PLOS Medicine, 2023. **20**(4): p. e1004215.

422. Rakotosamimanana, S., et al., *Spatial Analysis of Pulmonary Tuberculosis in Antananarivo Madagascar: Tuberculosis-Related Knowledge, Attitude and Practice.* PLOS ONE, 2014. **9**(11): p. e110471.

423. Polling, C., et al., *Understanding geographical patterning of self-harm prevalence within a diverse urban population: a mixed methods spatial analysis and qualitative study.* The Lancet (British edition), 2019. **394**: p. S78-S78.

424. Sangal, V. and P.A. Hoskisson, *Evolution, epidemiology and diversity of Corynebacterium diphtheriae: New perspectives on an old foe.* Infection, genetics and evolution, 2016. **43**: p. 364-370.

425. Seth-Smith, H.M. and A. Egli, *Whole genome sequencing for surveillance of diphtheria in low incidence settings.* Frontiers in public health, 2019. **7**: p. 235.

426. Howard, G. and V.J. Howard, *Design and implementation of observational studies to measure disease burden with a focus on stroke.* International Journal of Stroke, 2017. **13**(2): p. 157-165.

427. Kitamura, N., et al., *Waning rate of immunity and duration of protective immunity against diphtheria toxoid as a function of age and number of doses: Systematic review and quantitative data analysis.* Human vaccines & immunotherapeutics, 2022. **18**(6): p. 2099700-2099700.

428. Hanckel, B., et al., *The use of Qualitative Comparative Analysis (QCA) to address causality in complex systems: a systematic review of research on public health interventions.* BMC Public Health, 2021. **21**(1): p. 877.

429. Im, H. and J. Huh, *Does Health Information in Mass Media Help or Hurt Patients? Investigation of Potential Negative Influence of Mass Media Health Information on Patients' Beliefs and Medication Regimen Adherence.* Journal of health communication, 2017. **22**(3): p. 214-222.

430. McCombs, M.E. and D.L. Shaw, *The Agenda-Setting Function of Mass Media*. 2008, London: SAGE Publications Ltd: London. p. II290.

431. Prygiel, M., et al., *Challenges of Diphtheria Toxin Detection.* Toxins, 2024. **16**(6): p. 245.

432. Rajamani Sekar, S.K., et al., *Strengthening the laboratory diagnosis of pathogenic Corynebacterium species in the Vaccine era.* Letters in Applied Microbiology, 2017. **65**(5): p. 354-365.

433. Higgins, J. and D.G. Altman, *Assessing risk of bias in included studies.* 2008.

434. Booth, A., *Systematic approaches to a successful literature review / Andrew Booth, Anthea Sutton and Diana Papaioannou*. 2nd ed, ed. A. Sutton and D. Papaioannou. 2016, London: London : Sage.

435. Mullen, P.D. and G. Ramírez, *The promise and pitfalls of systematic reviews.* Annual review of public health, 2006. **27**(1): p. 81-102.

Appendices

Appendix 1. University of Nottingham's School of Medicine Research Ethics Committee approval for the qualitative study

University of Nottingham
UK | CHINA | MALAYSIA

Faculty of Medicine & Health Sciences
Research Ethics Committee
Faculty Hub
Room E41, E Floor, Medical School
Queen's Medical Centre Campus
Nottingham University Hospitals
Nottingham, NG7 2UH
Email: FMHS-ResearchEthics@nottingham.ac.uk

07 March 2022

Juniorcaius Ikejezie
PhD Student, Epidemiology and Public Health
Population and Lifespan Sciences
School of Medicine
Room B126, Clinical Sciences Building
City Hospital Campus, Nottingham University Hospitals
Nottingham, NG5 1PB

Dear Mr Ikejezie

Ethics Reference No: FMHS 454-0122 **– please always quote**	
Study Title: Prevention and Control of Diphtheria in Haiti: A qualitative Study of Stakeholders' Perspectives	
Location of Study: Port-au-Prince, Haiti	
Chief Investigator/Supervisor: Revati Phalkey, Associate Professor, Epidemiology and Public Health, Population and Lifespan Sciences, School of Medicine	
Lead Investigators/student: Juniorcaius Ikejezie, PhD student, Epidemiology and Public Health, Population and Lifespan Sciences, School of Medicine	
Other Key Investigators/Collaborators: Tessa Langley, Sarah Lewis, Professor of Medical Statistics, Epidemiology and Public Health, Population and Lifespan Sciences, School of Medicine, Dr Mariette Bermuda Francois Petit, Haiti's Ministry of Public Health and Population, Directorate of Epidemiology, Laboratory and Research (DELR), Haiti	
Proposed Start Date: 01/03/2022	**Proposed End Date:** 30/11/2022

The Committee considered this application at its meeting on 11 February 2022 and the following documents were received:

- FMHS REC Application form and supporting documents version 1.0: 24.01.2022

These have been reviewed and the research project is given a favourable ethics opinion.

Please submit a copy of the letter of approval from the Haiti's National Committee of Bioethics (NCB) so it can be noted to file. This must be completed before recruitment of participants commences.

Please note that this favourable ethics opinion is given on the understanding that:

1. All appropriate ethical requirements and regulatory permissions are respected and followed in accordance with all local laws of the country in which the study is being conducted and those required by the host organisation/s involved.
2. The research project will adhere to ICH E6 (R2) Good Clinical Practice (GCP) which is the international ethical, scientific, and practical standard to which all clinical research is conducted.
3. The protocol agreed is followed and the Committee is informed of any changes using a notice of amendment form (please request a form).
4. The Chair is informed of any serious or unexpected adverse event.
5. An End of Project Progress Report is completed and submitted to FMHS REC within six months after the study has finished (Please request a form).

Yours sincerely

Dr John Williams, Associate Professor
Chair, Faculty of Medicine & Health Sciences Research Ethics Committee

Appendix 2. University of Nottingham's School of Medicine Research Ethics Committee approval for the X study

University of Nottingham
UK | CHINA | MALAYSIA

Faculty of Medicine & Health Sciences
Research Ethics Committee
Faculty Hub
Room E41, E Floor, Medical School
Queen's Medical Centre Campus
Nottingham University Hospitals
Nottingham, NG7 2UH
Email: FMHS-ResearchEthics@nottingham.ac.uk

10 August 2022

Juniorcaius Ikejezie
PhD Student, Epidemiology and Public Health
Population and Lifespan Sciences
School of Medicine
Clinical Sciences Building
City Hospital Nottingham Campus
Hucknall Road
Nottingham
NG5 1PB

Dear Mr Ikejezie

Ethics Reference No: FMHS 38-0722– **please always quote**	
Study Title: Public perceptions of diphtheria in Haiti, December 2014-December 2021: a mixed-methods study using Twitter data.	
Chief Investigator/Supervisor: Sarah Lewis, Professor of Medical Statistics, Epidemiology and Public Health, Population and Lifespan Sciences, School of Medicine	
Lead Investigators/student: Juniorcaius Ikejezie, PhD student, Epidemiology and Public Health, School of Medicine.	
Other Key investigators: Ravati Phalkey, Honorary Associate Professor, Climate Change and Health in the UK, Epidemiology and Public Health, School of Medicine, Tessa Langley, Associate Professor in Health Economics, Epidemiology and Public Health, School of Medicine, Donal Bisanzio, Honorary Assistant Professor in infectious diseases in humans and animals, Epidemiology and Public Health, School of Medicine	
Proposed Start Date: 01/07/2022	**Proposed End Date:** 31/12/2022

Thank you for submitting the above application which was considered by a sub-committee meeting on 20 July 2022 and the following documents were received:

- FMHS REC Application form and supporting documents version 1.0: 04/07/2022

These have been reviewed and are satisfactory and the study is given a favourable ethics opinion.

A favourable research ethics opinion is given on the understanding that:

1. Confirmation from Twitter that the study adheres to the Twitter Developer Agreement and Policy is submitted when available and before commencement of the study,
2. The protocol agreed is followed and the Committee is informed of any changes using a notice of amendment form (please request a form).
3. The Chair is informed of any serious or unexpected event.
4. An End of Project Progress Report is completed and returned when the study has finished (Please request a form).

Yours sincerely

pp

Bethan E Phillips, Professor of Translational Physiology, School of Medicine
Centre Of Metabolism, Ageing & Physiology (COMAP) Injury, Inflammation & Recovery Sciences
Acting Chair, Faculty of Medicine & Health Sciences Research Ethics Committee

Appendix 3: PRISMA checklist

Section / Topic	Item No.	Checklist item	Reported on page No.
Title			
Title	1	Identify the report as a systematic review, meta-analysis, or both	1
Abstract			
Structured summary	2	Provide a structured summary including, as applicable, background, objectives, data sources, study eligibility criteria, participants, interventions, study appraisal and synthesis methods, results, limitations, conclusions and implications of key findings, systematic review registration number	N/A
Introduction			
Rationale	3	Describe the rationale for the review in the context of what is already known	1-2
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS)	2
Methods			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (such as web address), and, if available, provide registration information including registration number	2
Eligibility criteria	6	Specify study characteristics (such as PICOS, length of follow-up) and report characteristics (such as years considered, language, publication status) used as criteria for eligibility, giving rationale	3
Information sources	7	Describe all information sources (such as databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched	2
Search	8	Present full electronic search strategy for at least one database, including any limits	43-47

		used, such that it could be repeated	
Study selection	9	State the process for selecting studies (that is, screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis)	3-4
Data collection process	10	Describe method of data extraction from reports (such as piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators	4
Data items	11	List and define all variables for which data were sought (such as PICOS, funding sources) and any assumptions and simplifications made	4
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis	4-5
Summary measures	13	State the principal summary measures (such as risk ratio, difference in means)	3
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (such as I^2 statistic) for each meta-analysis	6
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (such as publication bias, selective reporting within studies)	5
Additional analyses	16	Describe methods of additional analyses (such as sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified	4
Results			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram	6-7
Study characteristics	18	For each study, present characteristics for which data were extracted (such as study size, PICOS, follow-up period) and provide	7-11

		the citations	
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome-level assessment (see item 12).	9-11, 45-47
Results of individual studies	20	For all outcomes considered (benefits or harms), present for each study (a) simple summary data for each intervention group and (b) effect estimates and confidence intervals, ideally with a forest plot	12-19
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency	15
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see item 15)	12-13, 20, 48-53
Additional analysis	23	Give results of additional analyses, if done (such as sensitivity or subgroup analyses, meta-regression) (see item 16)	54-55
Discussion			
Summary of evidence	24	Summarise the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (such as healthcare providers, users, and policy makers)	20-26
Limitations	25	Discuss limitations at study and outcome level (such as risk of bias), and at review level (such as incomplete retrieval of identified research, reporting bias)	26-27
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research	28
Funding			
Funding	27	Describe sources of funding for the systematic review and other support (such as supply of data) and role of funders for the systematic review	2

Appendix 4: Systematic review search strategy

EMBASE and MEDLINE

#	Search terms
1	exp Risk Factors/
2	exp Epidemiologic Studies/
3	exp Odds Ratio/
4	exp Multivariate Analysis/
5	exp Cross-Sectional Studies/
6	exp Case-Control Studies/
7	exp Cohort Studies/
8	exp Longitudinal Studies/
9	exp Prospective Studies/
10	exp Retrospective Studies/
11	exp Logistic Models/
12	exp Prevalence/
13	exp Incidence/
14	exp Probability/
15	epidemiolog$.ti,ab
16	ecologic$.ti,ab
17	factor$.ti,ab
18	probabilit$.ti,ab
19	frequenc$.ti,ab
20	characteristic$.ti,ab
21	criteri$.ti,ab
22	prognos$.ti,ab
23	(case-control or cohort or cross-sectional or ecologic$ or retrospective or prospective or longitudinal stud$).ti,ab.
24	risk$.ti,ab.
25	predict$.ti,ab.
26	correlat$.ti,ab.
27	etiol$.ti,ab.
28	aetiol$.ti,ab.
29	prevalence$.ti,ab.
30	incidence$.ti,ab.
31	rate$.ti,ab.

32	determinant$.ti,ab.
33	odds ratio$.ti,ab.
34	associat$.ti,ab.
35	stratif$.ti,ab
36	(ROC Curve$).ti,ab.
37	discriminat$.ti,ab
38	c statistic$.ti,ab
39	(Area under the curve$).ti,ab.
40	AUC$.ti,ab
41	outcome$.ti,ab
42	indice$.ti,ab
43	algorithm$.ti,ab
44	multivaria$.ti,ab
45	model$.ti,ab
46	or/1-45
47	diphtheria$.ti,ab.
48	exp Diphtheria/
49	47 or 48
50	46 and 49

PubMed and Web of Science

#	Search terms
1	(diphtheria OR diphtheriae) AND (epidemiolog* OR ecologic* OR factor* OR probabilit* OR frequenc* OR characteristic* OR criteri* OR prognos* case-control OR cohort* OR cross-sectional OR ecologic* OR retrospective OR prospective OR longitudinal OR risk* OR predict* OR correlat* OR etiol* OR aetiol* OR prevalence* OR incidence* OR rate* OR determinant* OR odds ratio* OR associat* OR stratif* OR ROC Curve* OR discriminat* OR c statistic* OR area under the curve* OR AUC OR outcome OR indice* OR algorithm OR multivaria* OR model*)

Appendix 5. Newcastle-Ottawa Scale assessment scores

Case-control studies

Reference	Selection				Comparability	Exposure			Score
	Adequate case definition	Cases were representative	Selection of community controls	Controls had no history of diphtheria	Comparability of cases and controls based on design or analysis	Ascertainment of exposure using a secure record or adequately blinded interviews	Same method of ascertainment for cases and controls	Same response rate for both groups	
Allam 2016	✓	✓		✓	✓✓		✓	✓	7
Bisgard 2000	✓	✓	✓	✓	✓✓	✓		✓	8
Bitragunta 2008	✓	✓	✓		✓✓		✓	✓	7
Brennan 2000	✓	✓	✓	✓	✓✓	✓	✓	✓	9
Chen 2000		✓	✓		✓✓	✓	✓	✓	7
Faria 1971					✓✓		✓	✓	4
Husada 2018	✓	✓			✓✓		✓	✓	6
Jones 1985	✓		✓		✓✓		✓	✓	6
Murakami 2010	✓	✓	✓	✓	✓✓		✓		7
Nassar 2021	✓	✓	✓	✓	✓✓		✓	✓	8
Quick 2000	✓	✓			✓✓		✓	✓	6
Ramdan 2018		✓	✓	✓	✓✓		✓	✓	7
Sein 2016	✓	✓	✓	✓	✓✓		✓	✓	8
Vitek 1999	✓	✓	✓		✓✓	✓	✓	✓	8

Cohort study

Reference	Selection				Comparability	Outcome			Score
	Exposed cohort was representative	Non-exposed cohort was drawn from the same community as the exposed cohort	Exposure was ascertained using secure record or structured interview	There was evidence that diphtheria was absent at the start of study	The study controlled for age, location, or other factors	Outcome was assessed through an independent blind assessment or record linkage	Follow-up was sufficiently long	Follow-up was ≥70%	
Chandra 1973		✓			✓✓	✓	✓	✓	6

Cross-sectional studies

Reference	Selection				Comparability	Outcome		Score
	Sample was representative	Sample size was justified	Response rate was satisfactory	Ascertainment of exposure using a validated tool	Outcome groups were comparable based on design or analysis	Outcome was assessed through an independent blind assessment or record linkage	Statistical test was described and appropriate	
Belsey 1969				✓✓		✓✓		4
Harnisch 1989	✓	✓	✓	✓✓	✓	✓✓		8
Kalapothaki 1984	✓		✓	✓✓	✓	✓✓		7
Kitamura 2023	✓	✓	✓	✓✓	✓	✓✓	✓	9
Marcuse 1973	✓	✓	✓	✓✓	✓	✓✓		8
Miller 1972	✓	✓	✓	✓✓	✓	✓✓		8
Ohuabunwo 2005	✓	✓	✓	✓✓	✓	✓✓	✓	9
Trichopoulos 1972	✓	✓	✓	✓✓	✓	✓✓		8

Appendix 6. GRADE assessment scores

Theme and risk factor	No.	Study design	Risk of bias	Imprecision	Inconsistency	Indirectness	Publication bias	Other considerations	Quality	Ref.
Vaccination or contact with cases										
Having a sibling with no basic immunization	1	Observational	Not serious	Serious	Not serious	Not serious	Not serious		Very low	[169]
Incomplete vaccination (<3 doses)	18	Observational	Not serious	Not serious	Not serious	Not serious	Not serious	Large effect	Moderate	[165-168, 170-176, 179, 181-186]
No booster vaccination in last five years	4	Observational	Not serious	Not serious	Not serious	Not serious	Not serious		Very low	[167, 172, 177, 185]
No DTP dose in last ten years	1	Observational	Not serious	Serious	Not serious	Not serious	Not serious		Very low	[182]
Having contact with a diphtheria case	5	Observational	Not serious	Not serious	Not serious	Not serious	Not serious		Low	[171-175]
Having contact with a person with skin lesions	3	Observational	Not serious	Not serious	Not serious	Not serious	Not serious		Low	[171, 172, 174]
Underlying conditions										
Having a history of a chronic condition	2	Observational	Not serious	Serious	Serious	Not serious	Not serious		Very low	[172, 174]

Having a history of eczema	1	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[174]
Having a recent sore throat	2	Observational	Not serious	Serious	Serious	Not serious	Not serious	Very low	[174, 186]
Having adenectomy (i.e., surgical removal of all or part of a gland)	1	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[172]
Having cough or rhinorrhea four weeks prior to diphtheria onset	1	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[174]
Having fever with myalgia four weeks prior to diphtheria onset	1	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[174]
Having skin lesions or rash four weeks prior to diphtheria onset	1	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[174]
Having tonsils	2	Observational	Not serious	Not serious	Serious	Not serious	Not serious	Very low	[174, 186]
Malnutrition	2	Observational	Not serious	Not serious	Not serious	Not serious	Not serious	Very low	[175, 182]

Knowledge and behaviour

Abusing alcohol	2	Observational	Not serious	Serious	Serious	Not serious	Not serious	Very low	[174, 180]
Attending social gatherings	2	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[174, 185]
Belief that vaccines are ineffective	1	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[164]
Consumption of cheese, ice cream, or raw milk	1	Observational	Not serious	Serious	Not serious	Serious	Not serious	Very low	[171]
Consumption of factory-made yoghurt	1	Observational	Not serious	Serious	Not serious	Serious	Not serious	Very low	[171]
Drinking directly from faucets or a common water container	1	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[185]
Having a guest spending a night in the house in last two weeks	1	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[174]
Having fears or worries about vaccines and their safety	2	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[164, 175]

Having low knowledge of diphtheria	3	Observational	Not serious	Not serious	Not serious	Not serious	Not serious	Low	[164, 170, 175]
Having no knowledge of diphtheria vaccine boosters	1	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[164]
Household member sought medical care in last two weeks	1	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[174]
Household member spent night outside home in last two weeks	1	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[174]
Infrequent bathing	3	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[172, 174, 182]
Kissing someone outside family in last two weeks	1	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[174]
Obtaining water from wheeled carrier	1	Observational	Not serious	Serious	Not serious	Not serious	Serious	Very low	[171]
Infrequent hand washing	2	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[172, 182]

Sharing a bed or bedroom	4	Observational	Not serious	Not serious	Serious	Not serious	Not serious	Very low	[173, 174, 176, 182]
Sharing body and/or hand towels	2	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[173, 174]
Sharing cigarettes	1	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[185]
Sharing toothbrushes	1	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[172]
Sharing utensils, cups, glasses	4	Observational	Not serious	Not serious	Serious	Not serious	Not serious	Very low	[172-174, 185]
Smoking cigarettes	1	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[174]
Staying in a dormitory	1	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[182]
Talking or shaking hands with people at work	1	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[174]
Travel history to area with diphtheria	2	Observational	Not serious	Not serious	Not serious	Not serious	Not serious	Very low	[173, 175]
Wearing someone else's clothes before washing	1	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[174]
Wearing the same clothes for multiple days prior to washing	1	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[174]

Working in a dusty, noisy or smoky environment	1	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[174]
Socioeconomic status									
Area of residence (urban or rural)	1	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[174]
Attending school or kindergarten	2	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[172, 182]
Being a refugee	1	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[174]
Being part of a large family or household	5	Observational	Not serious	Not serious	Not serious	Not serious	Not serious	Low	[164, 170, 173, 174, 178]
Being unemployed	2	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[164, 174]
Having a caregiver with low or no education	1	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[164]
Having a father with low education	2	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[170, 173]
Having a low income	2	Observational	Not serious	Serious	Not serious	Serious	Not serious	Very low	[164, 178]
Having a mother with low education	3	Observational	Not serious	Serious	Serious	Not serious	Not serious	Very low	[170, 173, 174]

Having an unemployed caregiver	2	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[164, 173]
Having contact with farm or pet animals in last two weeks	2	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[174, 182]
Having low or no education	2	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[164, 174]
Having no access to clean water at home	1	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[170]
Having no standard toilet and washing facilities at home	1	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[170]
Having poor ventilation at home	2	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[170, 175]
Having several siblings	1	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[176]
Living in a house with few rooms	4	Observational	Not serious	Not serious	Not serious	Not serious	Not serious	Low	[164, 172-174]
Living in a house with several children	2	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[174, 176]

Living in an overcrowded house	4	Observational	Not serious	Serious	Not serious	Serious	Not serious	Very low	[170, 173, 175, 178]
Residential status (permanent or temporary)	1	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[172]
Size of the household	3	Observational	Not serious	Not serious	Not serious	Not serious	Not serious	Very low	[170, 178, 182]
Type of home floor	1	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[175]
Type of home wall	1	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[175]
Type of house	3	Observational	Not serious	Serious	Not serious	Not serious	Not serious	Very low	[172-174]
Population-level factors									
Adult morbidity	1	Observational	Serious	Serious	Not serious	Not serious	Not serious	Very low	[191]
Conflict in past year	1	Observational	Serious	Serious	Not serious	Not serious	Not serious	Very low	[188]
Degree of urbanization	2	Observational	Serious	Serious	Serious	Not serious	Not serious	Very low	[189, 191]
Diphtheria tetanus (DT) vaccine stockout	1	Observational	Serious	Serious	Not serious	Not serious	Not serious	Very low	[189]
DTP vaccine stockout	1	Observational	Serious	Serious	Not serious	Not serious	Not serious	Very low	[189]
Female literacy	1	Observational	Serious	Serious	Not serious	Not serious	Not serious	Very low	[189]
Health facility density	1	Observational	Serious	Serious	Not serious	Not serious	Not serious	Very low	[189]
Level of education	3	Observational	Serious	Not serious	Serious	Not serious	Not serious	Very low	[190-192]

Level of emissions into the air from stationary or mobile sources	1	Observational	Serious	Serious	Not serious	Not serious	Not serious	Very low	[191]
Level of sulfur dioxide in the air	1	Observational	Serious	Serious	Not serious	Not serious	Not serious	Very low	[191]
Levels of carbon monoxide, dust, or nitrogen dioxide in the air	1	Observational	Serious	Serious	Not serious	Not serious	Not serious	Very low	[191]
Median income	1	Observational	Serious	Serious	Not serious	Not serious	Not serious	Very low	[192]
Migration balance	1	Observational	Serious	Serious	Not serious	Not serious	Not serious	Very low	[191]
Morbidity in the rural population	1	Observational	Serious	Serious	Not serious	Not serious	Not serious	Very low	[191]
Morbidity in the urban population	1	Observational	Serious	Serious	Not serious	Not serious	Not serious	Very low	[191]
Mortality in the overall population	1	Observational	Serious	Serious	Not serious	Not serious	Not serious	Very low	[191]
Number of immunized children in areas with past conflict	1	Observational	Serious	Serious	Not serious	Not serious	Not serious	Very low	[188]

Number of immunized children in areas with ongoing conflict	1	Observational	Serious	Serious	Not serious	Not serious	Not serious	Very low	[188]
Ongoing conflict	1	Observational	Serious	Serious	Not serious	Not serious	Not serious	Very low	[188]
Percentage of households with no toilet facility	1	Observational	Serious	Serious	Not serious	Not serious	Not serious	Very low	[189]
Percentage of people below poverty line	1	Observational	Serious	Serious	Not serious	Not serious	Not serious	Very low	[191]
Percentage of overcrowded housing units	1	Observational	Serious	Serious	Not serious	Not serious	Not serious	Very low	[192]
Population expenditure per capita	1	Observational	Serious	Serious	Not serious	Not serious	Not serious	Very low	[190]
Population growth rate	1	Observational	Serious	Serious	Not serious	Not serious	Not serious	Very low	[191]
Population density	4	Observational	Serious	Not serious	Not serious	Not serious	Not serious	Very low	[189-192]
School density	1	Observational	Serious	Serious	Not serious	Not serious	Not serious	Very low	[189]
Tuberculosis incidence	1	Observational	Serious	Serious	Not serious	Not serious	Not serious	Very low	[187]
Vaccination coverage	4	Observational	Serious	Not serious	Not serious	Not serious	Not serious	Very low	[188-191]

Appendix 7. Reported risk estimates

The tables below present all the risk estimates from each study included in the systematic review.

Case-control studies

Allam 2016

	Univariate		Multivariate	
Variables	**cOR**	**95% CI**	**aOR**	**95% CI**
Education	1.45	0.54–3.90		
Education of caretaker	1.07	0.52–2.15		
Occupation	1	0.48–2.05		
Occupation of caretaker	0.51	0.21–1.20		
Eligible for white ration card	0.68	0.33–1.37		
Mean persons sleeping in house	0.68	0.31–1.47		
Mean rooms in house	0.87	0.42–1.80		
Heard of diphtheria	3.73	1.69–8.17	3.56	1.58–8.04
Knowledge about booster doses of diphtheria				
Additional (booster) doses of diphtheria vaccine	0.22	0.10–0.51		
Booster doses are needed to stay protected	0.21	0.09–0.48		
How many booster doses are recommended	0.10	0.01–0.79		
What age a child should get the final booster dose	0.12	0.03–0.42		
Attitude toward vaccines				
Vaccines can prevent people from getting diseases	4.61	1.44–14.8	3.99	1.18–13.45
Vaccines are safe	1.34	0.29–6.32		
Fear or worries about vaccines	1.93	0.61–6.13		

aOR, adjusted odds ratio; CI, confidence interval; cOR, crude odds ratio.

Bisgard 2000

Variables	**Cases**	**Controls**	**Cross-tabulated OR (95% CI)**
Vaccine doses			
<3	61	171	4.57 (3.27–6.38)
≥3	156	1997	

OR, odds ratio; CI, confidence interval; N/A, not available.

Bitragunta 2008

Variables	**Cases**	**Controls**	**Cross-tabulated OR (95% CI)**
Vaccine doses			
<3	41	18	2.92 (1.56–5.45)
≥3	82	105	

OR, odds ratio; CI, confidence interval; N/A, not available.

Brennan 2000

Variables	Cases	Controls	OR (95% CI)
Vaccine doses			
<3	34	84	2.67 (0.96–7.42) #
≥3	5	33	
Years since last dose			
≥5	6	2	12.7 (1.5–106.6)
0-4	19	50	

OR, odds ratio; CI, confidence interval; N/A, not available. # Calculated using 2x2 cross-tabulation.

Chandra 1973

Variables	Carriers	Controls	Cross-tabulated OR (95% CI)
Social class			
V	4	40	0.71 (0.21–2.41)
IV	10	71	
Overcrowding			
Present	3	53	0.31 (0.08–1.19)
Absent	11	61	
Members in the family			
≥7	5	57	0.56 (0.18–1.76)
<7	9	57	

OR, odds ratio; CI, confidence interval; N/A, not available.

Chen 2000

Variables	Cases	Controls	Cross-tabulated OR (95% CI)
Vaccine doses			
<3	87	29	8.37 (5.31–13.18) #
≥3	175	488	

OR, odds ratio; CI, confidence interval; N/A, not available.

Faria 1971

Variables	Cases	Controls	Cross-tabulated OR (95% CI)
Vaccine doses			
<3	89	59	4.07 (1.84–9.03)
≥3	10	27	

OR, odds ratio; CI, confidence interval.

Husada 2018

	Univariate		Multivariate	
Variables	**cOR**	**95% CI**	**aOR**	**95% CI**
Paternal education	5.50	1.560–19.392	5.5	1.560–19.392
Maternal education	3.077	0.862–10.978	0.921	0.169–5.009
Knowledge about diphtheria	3.836	1.081–13.604	2.434	0.647–9.159
Immunization: complete basic	1.517	0.649–3.547		
District	1.563	0.328–7.438		
Living in "pesantren"	0.481	0.058–4.018		
No. of people living in house (>2 vs <2 people)	1.511	0.607–3.763		
Ventilation	2.669	0.582–12.239		
Clean water source	2.0	0.826–4.844	1.161	0.448–3.007
Toilet and washing facilities	1.217	0.826–4.844		

aOR, adjusted odds ratio; CI, confidence interval; cOR, crude odds ratio.

Jones 1985

	n=46		Univariate		Multivariate	
Variables	**Cases (%)**	**Controls (%)**	**RR**	**95% CI**	**RR**	**p value**
Vaccine doses						
<3	89	67	3.97 #	1.30–12.09		
≥3	11	33				
Contact with a diphtheria case	43	6	31.65	4.22–237.5	71.04	0.002
Contact with skin disease	22	7	3.84	1.16–12.7	41.91	0.04
Obtaining drinking-water from wheeled carrier	26	10	10.28	1.24–85.21	28.42	0.008
Milk and milk products						
Raw milk	17	7	4.25	0.83–21.9		
Factory-made yoghurt	60	24	3.88	1.71–8.8	14.92	0.003
Cheese	85	71	2.08	0.69–6.25		
Ice-cream	51	36	2.15	0.97–4.74		

RR, relative risk; CI, confidence interval; N/A, not available.
Calculated using 2x2 cross-tabulation.

Murakami 2010

Variables	**Cases**	**Controls**	**Cross-tabulated OR (95% CI)**	
Vaccine doses				
<3	10	26	1.61 (0.74–3.47)	
≥3	78	326		
	Univariate		**Multivariate**	
	cOR	**95% CI**	**aOR**	**95% CI**
Bathing frequency				
Once a day or less	1.73	1.05–2.86	1.74	1.04–2.91
Twice a day or more	1.00			

aOR, adjusted odds ratio; CI, confidence interval; cOR, crude odds ratio.

	Univariate		Multivariate	
Variables	**cOR**	**95% CI**	**aOR**	**95% CI**
Paternal factors				
Low education level	0.8	0.4–1.4	0.6	0.3–1.5
Not working	1.3	0.6–2.8	1.1	0.5–2.5
Maternal factors				
Low education level	0.9	0.4–1.9	0.9	0.4–2.5
Not working	1.0	0.4–2.5	0.9	0.3–2.7
Household factors				
Type of house (apartment)	1.1	0.6–1.9	0.9	0.4–1.8
Family size of at least seven people (median)	0.7	0.4–1.3	0.6	0.3–1.2
Three or more rooms in house (mean)	0.9	0.5–1.6	0.8	0.4–1.7
Shared a bedroom with at least two people (mean)	2.1	1.0–4.2	2.8	1.2–6.6
Shared utensils and cups	1.2	0.4–3.2	1.1	0.3–3.3
Shared towels	1.3	0.7–2.5	0.9	0.4–1.9
Immunization status				
Non-vaccinated	2.3	1.2–4.6	2.6	1.2–6.0
Source of infection				
Travel history to area with diphtheria	1.2	0.4–3.2	0.5	0.2–1.7
Contact with a diphtheria case	9.1	2.6–31.5	10.6	2.6–43.6

aOR, adjusted odds ratio; CI, confidence interval; cOR, crude odds ratio.

Variables	Matched OR	95% CI
Not vaccinated	19.23	4.79–77.17
Exposure to infected person		
Household member with diphtheria	7.46	1.46–37.99
Exposed to others with skin lesions	5.86	1.92–17.84
Host factors		
Person with tonsils	4.41	2.03–9.56
Fever and myalgia 4 weeks prior to reference date	2.67	1.29–5.52
History of chronic health problem	2.12	1.20–3.76
History of eczema	3.41	1.17–9.92
Bathed less than once a week	2.60	1.29–5.23
Wore someone else's clothes before washing them	0.7	0.5–1.1
Wore clothes >3 days before washing	1.1	0.7–1.6
Shared body and hand towel with family members	1.1	0.7–1.8
Shared hand towel at work	1.6	0.5–5.4
Shared hand towel with family members	1.2	0.8–1.8
Shared utensils outside family	1.3	0.4–4.3
Kissed someone outside of family in last 2 weeks	1.0	0.6–1.7
Kissed someone outside of family on lips in last 2 weeks	1.3	0.8–2.3
Shook hands on daily basis with the public at work	1.0	0.3–2.9
Talked or shook hands with 110 coworkers (contact within <1 m)	0.6	0.2–1.7
Nonsmokers	1.61	1.00–2.63
Heavy drinker (>14 drinks/week)	0.7	0.3–1.7
Being a refugee	0.9	0.2–2.7
Exposure to animals		
Touched farm animals in last 2 weeks	0.7	0.5–1.1
Milked a cow or goat in last 2 weeks	0.9	0.5–1.3
Pets living in home	1.0	0.7–1.4
Household characteristics		
Shared utensils, cups, or glasses with family	2.72	1.58–4.68
Shared a bed (with ≥1 person)	1.96	1.13–3.40
Apartment vs. house	0.7	0.4–1.1
No. of rooms in home (1–4 vs. 14)	1.7	1.0–2.8
No. of persons in household (>5)	1.0	0.7–1.4
Guest spent night in household 2 weeks prior to diphtheria onset	1.3	0.96–1.94
Work environment		
Dusty	1.8	0.5–5.4
Very noisy	1.4	0.5–3.6
Heavy cigarette smoke	1.9	0.7–5.2
Exposure outside home 2 weeks before reference date		
Attended church	0.6	0.3–1.2
Attended sports club	0.4	0.1–1.2
Attended social clubs	0.2	0.07–1.2
Attended movie	0.9	0.5–1.6
Attended funeral	0.7	0.5–1.2
Attended celebration	0.9	0.6–1.3
Cough, rhinorrhea, or sore throat 4 weeks before reference date	1.3	0.94–1.8

OR, odds ratio; CI, confidence interval.

Ramdan 2018

	Bivariate		Multivariate	
Variables	**cOR**	**95% CI**	**aOR**	**95% CI**
DTP immunization status (i.e., having received three vaccine doses)	0.343	0.073–1.1.617		
Nutritional status	4.457	1.11–17.89	0.810	0.065–10.073
Mobility (i.e., travel to an area with diphtheria cases)	6.800	2.253–31.645	8.456	5.643–12.672
Source of contamination (i.e., contact with diphtheria case)	0.167	0.039–0.711	0.134	0.012–1.519
Knowledge of parent	1.111	0.306–4.037		
Attitude towards immunization program	0.889	0.244–3.243		
Home ventilation	1.615	0.370–7.049		
Home density of occupancy	0.833	0.203–3.427		
Home wall type	1.143	0.307–4.254		
Home floor type	1.167	0.302–4.512		

aOR, adjusted odds ratio; CI, confidence interval; cOR, crude odds ratio.

Sein 2016

Variables	**Cases**	**Controls**	**OR (95% CI)**
Mother's occupation			
Farmer	35	77	0.00 (0.00–∞ i.e., infinity)
Non-farmer	3	2	
Number of children <5 years in household			
>2	26	60	1.43 (0.55–3.73)
≤2	11	19	
Number of persons in household sharing bed			
>4	23	46	0.39 (0.15–1.07)
≤4	15	33	
Travel within 2 weeks			
Yes	2	2	2.72 (0.23–32.78)
No	36	76	
Chronic skin condition / lesion			
Yes	0	0	N/A
No	37	79	
Vaccine doses			
<3	30	57	2.11 (0.72–6.17) #
≥3	5	20	

OR, odds ratio; CI, confidence interval; N/A, not available. # Calculated using 2x2 cross-tabulation.

Vitek 1999

	Univariate		Multivariate	
Variables	**cOR**	**95% CI**	**aOR**	**95% CI**
Number of doses				
4	2.8	1.2–6.5	1.7	0.5–5.0
5	1.0§		1.0§	
	Cases	**Controls**		**Cross-tabulated OR (95% CI)**
Years since last dose				
≥5	11	37		1.70 (0.81–3.57) #
<5	47	269		

aOR, adjusted odds ratio; CI, confidence interval; cOR, crude odds ratio. § Reference group.

Cross-sectional studies

Belsey 1969

Variables	**Carriers**	**Controls**	**Cross-tabulated OR (95% CI)**
Vaccine doses			
<3	10	68	2.51 (1.07–6.31)
≥3	10	171	

OR, odds ratio; CI, confidence interval.

Harnisch 1989

Variables	**Cases / Carriers**	**Controls**	**Cross-tabulated OR (95% CI)**
Alcohol abuse			
Yes	237	935	48.82 (27.20–87.61)
No	12	2,311	

OR, odds ratio; CI, confidence interval.

Kalapothaki 1984

Variables	**Carriers**	**Controls**	**Cross-tabulated OR (95% CI)**
Vaccine doses			
<3	1	205	0.56 (0.07–4.64)
≥3	6	683	

OR, odds ratio; CI, confidence interval.

Kitamura 2023

Variables	Carriers	Controls	p value	Cross-tabulated OR (95% CI)
DTP1 ≤10 years				
0 doses	5	107	>0.99	
≥1 dose	16	336		
DTP3 ≤10 years				
0 doses	5	111	>0.99	0.94 (0.34–2.61)
≥1 dose	16	332		
School				
Not attended	17	906	0.12	
Attended	6	283		
Dormitory				
Not staying	23	1,035	0.77	
Staying	4	154		
Sharing a bed				
Yes	4	143	0.67	1.26 (0.43–3.70)
No	23	1,037		
Household size, no. persons				
≤4	13	585	>0.99	
>4	14	604		
Bathing, times/day				
<1	0	72	0.40	
≥1	27	1,117		
Handwashing, times/day				
<3	4	247	0.11	
≥3	18	445		
Livestock or pet animal				
No	24	866	0.08	
Yes	3	323		

OR, odds ratio; CI, confidence interval.

Marcuse 1973

Variables	Cases / Carriers	Controls	Cross-tabulated OR (95% CI)
Vaccine doses			
<3	44	305	1.12 (0.76–1.64)
≥3	87	673	

OR, odds ratio; CI, confidence interval.

Miller 1972

Variables	Cases	Controls	Cross-tabulated OR (95% CI)
Vaccine doses			
<3	31	66	0.88 (0.52–1.46)
≥3	73	136	

OR, odds ratio; CI, confidence interval.

Ohuabunwo 2005

Variable	RR (95% CI)
Sharing cups	1.36 (1.06–1.74)

OR, odds ratio; CI, confidence interval.

Trichopoulos 1972

Variables	Carriers	Controls	Cross-tabulated OR (95% CI)
Vaccine doses			
<3	22	505	2.07 (0.83–5.16)
≥3	6	285	
Tonsillectomy			
Yes	4	99	1.16 (0.40–3.42)
No	24	691	

OR, odds ratio; CI, confidence interval.

Ecological studies

Coleman 2018

City	Tuberculosis coefficient	Standard error	p value	R^2	N
Boston 1916–1923	0.14	0.07	0.047	0.51	411
Chicago 1916–1923	0.11	0.03	0.0004	0.66	409
Detroit 1915–1923	0.12	0.05	0.01	0.68	453
New York 1916–1923	0.03	0.04	0.41	0.63	408
New York w/o 1921	0.07	0.04	0.05	0.64	357
Philadelphia 1907–1923	0.06	0.03	0.04	0.57	823

R^2, coefficient of determination.

Dureab 2018

	Bivariate		Multivariate	
Variables	cOR	95% CI	aOR	95% CI
No. of immunized children	1.02	1.01–1.03	1.04	1.01–1.058
Conflict in past year	0.59	0.33–1.04	3.25	0.16–67.70
Ongoing conflict	1.89	1.20–2.99	11.21	1.29–97.69
No. of immunized children in areas with past conflict	0.99	0.99–1.00	0.98	0.95–1.02
No. of immunized children in areas with ongoing conflict	1.01	1.00–1.02	0.98	0.95–1.01

aOR, adjusted odds ratio; CI, confidence interval; cOR, crude odds ratio.

Ikejezie 2022

Variables	Initial OLS (SE)	Final OLS (SE)	Final GWR
DTP3 coverage	0.177 (0.488)		
Health facility density	0.015 (0.007)*	0.015 (0.005)**	0.015
Improved water source	0.003 (0.003)		
Female literacy	-0.026 (0.007)***	-0.024 (0.006)***	-0.024
No toilet facility	< -0.001 (0.004)		
School density	-0.001 (0.002)		
Population density	< -0.001 (<0.001)		
Urbanization	0.007 (0.003)**	0.006 (0.002)**	0.006

OLS, ordinary least squares; GWR, geographically weighted regression; SE, standard error.
* p<0.05.
** p<0.01.
*** p<0.001.

Izza 2015

Variables	r	p value
Population density		
2010	0.002	0.991
2011	0.072	0.668
Education level		
2010	0.059	0.725
2011	0.089	0.089
DTP3 coverage		
2010	0.424	0.008
2011	0.221	0.183
DT coverage		
2010	0.348	0.032
2011	0.198	0.232
Healthy home		
2010	0.125	0.454
2011	0.104	0.232
Per capita expenditure		
2010	0.037	0.823
2011	0.065	0.700

r, Spearman's rank correlation coefficient.

Podavalenko 2018

Variables	Regression coefficient (B)	Standard error	χ^2 walda	p value	95% PI for odds ratio
Environmental factors					
Sulfur dioxide	0.226	0.055	16.7	<0.001	1.125–1.396
Social factors					
Vaccination	-0.043	0.017	6.6	0.001	0.926–0.990
Generalized environmental factors (indices of human development)					
Living conditions, incl. population density	0.037	0.009	18.0	<0.001	1.020–1.056
Expeditors of risk					
Morbidity rate of city residents	0.153	0.063	5.9	0.001	1.031–1.318

PI, prediction interval.

Quesada 1979

Variables	Regression coefficient (B)	p value	F
Percentage of people below poverty	0.023	<0.05	37.196
Population density	0.004	>0.05	0.002
Percentage living in overcrowded households	-0.092	>0.05	0.590
Median education	0.162	>0.05	1.046
Median income	0.074	>0.05	0.260

PI, prediction interval.

Appendix 8. Haiti's National Bioethics Committee approval for the spatial analysis

REPUBLIQUE D'HAITI

MINISTERE DE LA SANTE PUBLIQUE
ET DE LA POPULATION
Comité National de Bioéthique

15 juin 2021

Evaluation initiale

Juniorcaius Ikejezie

Réf: 1921-45 : ***Une analyse spatiale de l'épidémiologie de la diphtérie en Haïti, version 1***

Monsieur Ikejezie,

Le Comité National de Bioéthique a revu le dossier, approuve le protocole et les instruments et donne son approbation du 15 juin 2021 au 14 janvier 2022.

Le Comité vous rappelle qu'il vous faut
- soumettre pour approbation avant implémentation une copie de toute modification apportée au protocole,
- soumettre une copie des différents rapports, publications et présentations qui seront élaborés à partir de cette évaluation,
- présenter les résultats en Haïti.

Le Comité vous souhaite du succès dans la conduite de cette étude.

Gérald Lerebours,
Président

Comité National de Bioéthique
c/o Association Médicale Haitienne (AMH)
29, 1[e] avenue du Travail, Port-au-Prince

Appendix 9. University of Nottingham's School of Medicine Research Ethics Committee approval for the spatial analysis

University of Nottingham
UK | CHINA | MALAYSIA

Email: FMHS-ResearchEthics@nottingham.ac.uk

Faculty of Medicine & Health Sciences
Research Ethics Committee
c/o Faculty PVC Office
School of Medicine Education Centre
B Floor, Medical School
Queen's Medical Centre Campus
Nottingham University Hospitals
Nottingham, NG7 2UH

8 April 2019

Juniorcaius Ikejezie
PhD Student
c/o Room C118 Clinical Sciences Building
School of Medicine, Epidemiology & Public Health
City Hospital Campus
Nottingham University Hospitals
Hucknall Road
Nottingham
NG5 1PB

Dear Juniorcaius Ikejezie

Ethics Reference No: 267-1903 – **please always quote**	
Study Title: A Spatial analysis of the epidemiology of diphtheria in Haiti	
Location of Study: Epidemiology and Public Health using anonymised secondary datasets acquired from Haiti	
Chief Investigator/Supervisor: Dr Ravati Phalkey, Assistant Professor, Epidemiology and Public Health, School of Medicine	
Lead Investigators/student: Juniorcaius Ikejezie, PhD Student, Epidemiology and Public Health	
Proposed Start Date: 01/05/2019	**Proposed End Date:** 30/09/2021 29 mths

Thank you for submitting the above application which was considered by the Committee and the following documents were received:

- FMHS REC Application form and supporting documents version 1.0: 20/03/2019

These have been reviewed and are satisfactory and the study has been given a favourable opinion.

A favourable opinion has been given on the understanding that:

1. All appropriate ethical and regulatory permissions are respected and followed in accordance with all local laws of the country in which the study is being conducted and those required by the host organisation/s involved
2. The protocol agreed is followed and the Committee is informed of any changes using a notice of amendment form (please request a form).
3. The Chair is informed of any serious or unexpected event.
4. An End of Project Progress Report is completed and returned when the study has finished (Please request a form).

Yours sincerely

Professor Ravi Mahajan
Chair, Faculty of Medicine & Health Sciences Research Ethics Committee

Appendix 10. Areas with spatial dependence in Haiti identified in the LISA analysis

Type of spatial dependence	n	Department	n	Commune
High-high	9	Centre	4	Boucan Carre, Hinche, Lascahobas, Saut d'Eau
		Nord-Est	2	Caracol, Terrier Rouge
		Ouest	2	Arcahaie, Thomazeau
		Artibonite	1	Marmelade
Low-low	14	Grand'Anse	7	Abricots, Anse d'Hainault, Chambellan, Jérémie, Moron, Pestel, Roseaux
		Sud	4	Chardonnieres, Les Anglais, Maniche, Torbeck
		Nippes	2	Baraderes, Plaisance du Sud
		Nord-Ouest	1	Baie de Henne
Low-high	6	Nord	5	La Victoire, Limonde, Milot, Plaine du Nord, Saint-Raphael
		Nord-Est	1	Perches
High-low	1	Sud	1	Fonds des Negres

Appendix 11. Variables collected from the X posts

Acronym	Meaning
uid	User ID
tlt	Post latitude
tln	Post longitude
t	Post text
src	Source of the post (device used)
sn	Screen name
plt	Place latitude
pln	Place longitude
name	Username
loc	Location
lang	Language of the post
id	Post ID
flrs	Number of followers
flng	Number of X accounts that the account is following
descr	Description of the account
cr	Post timestamp
cc	Country code
acr	Account creation timestamp

Appendix 12. University of Nottingham's School of Medicine Research Ethics Committee approval of the amended protocol for the X study

University of Nottingham
UK | CHINA | MALAYSIA

Faculty of Medicine & Health Sciences
Research Ethics Committee
Faculty Hub
Room E41, E Floor, Medical School
Queen's Medical Centre Campus
Nottingham University Hospitals
Nottingham, NG7 2UH
Email: FMHS-ResearchEthics@nottingham.ac.uk

04 April 2023

Juniorcaius Ikejezie
PhD Student, Epidemiology and Public Health
Population and Lifespan Sciences
School of Medicine
Clinical Sciences Building
City Hospital Nottingham Campus
Hucknall Road
Nottingham
NG5 1PB

Dear Mr Ikejezie

Ethics Reference No: FMHS 38-0722**– please always quote**	
Study Title: Public perceptions of diphtheria in Haiti, December 2014-December 2021: a mixed-methods study using Twitter data.	
Chief Investigator/Supervisor: Sarah Lewis, Professor of Medical Statistics, Epidemiology and Public Health, Population and Lifespan Sciences, School of Medicine	
Lead Investigators/student: Juniorcaius Ikejezie, PhD student, Epidemiology and Public Health, School of Medicine.	
Other Key investigators: Ravati Phalkey, Honorary Associate Professor, Climate Change and Health in the UK, Epidemiology and Public Health, School of Medicine, Tessa Langley, Associate Professor in Health Economics, Epidemiology and Public Health, School of Medicine, Donal Bisanzio, Honorary Assistant Professor in infectious diseases in humans and animals, Epidemiology and Public Health, School of Medicine	
Proposed Start Date: 01/07/2022	**Proposed End Date:** 31/12/2023

Thank you for notifying the Committee of amendment no 1: 21.0.2023 in summary as follows:

- Extend the timeframe of data to be analysed from December 2014-December 2021 to December 2014-December 2022.
- Addition of use of data from a large dataset of researchers at Boston Children's Hospital, Harvard University

and the following revised documents were received:

- FMHS REC Application form and Protocol version 1.0: 04/07/2022
- Boston Children's Hospital Data Use agreement Healthmap 46 and University of Nottingham V 2 dated 31.03.2023

These have been reviewed and are satisfactory and the study amendment no 1: 21.03.2023 is given a favourable ethics opinion.

Yours sincerely

pp

Bethan E Phillips, Professor of Translational Physiology, School of Medicine
Centre Of Metabolism, Ageing & Physiology (COMAP) Injury, Inflammation & Recovery Sciences
Acting Chair, Faculty of Medicine & Health Sciences Research Ethics Committee

Appendix 13. Diphtheria-related posts by country and corresponding income level

#	Income level	WHO region	Country	n.	%
1	High	Americas	United States	1146	47.0
2	High	Europe	United Kingdom	491	20.1
3	High	Americas	Canada	178	7.3
4	High	Western Pacific	Australia	137	5.6
5	Low-middle	South-East Asia	India	67	2.7
6	Upper-middle	Africa	South Africa	41	1.7
7	High	Europe	Spain	32	1.3
8	High	Europe	Ireland	31	1.3
9	Low-middle	Western Pacific	Philippines	29	1.2
10	Upper-middle	Western Pacific	Malaysia	28	1.1
11	Low	Africa	Nigeria	18	0.7
12	Low-middle	Eastern Mediterranean	Pakistan	17	0.7
13	High	Western Pacific	New Zealand	13	0.5
14	High	Europe	Germany	12	0.5
15	Upper-middle	South-East Asia	Indonesia	12	0.5
16	Upper-middle	Americas	Mexico	12	0.5
17	High	Europe	France	10	0.4
18	High	Western Pacific	Japan	8	0.3
19	Low	Africa	Uganda	7	0.3
20	Low	Eastern Mediterranean	Yemen	7	0.3
21	Low-middle	Africa	Kenya	6	0.2
22	High	Europe	Netherlands	6	0.2
23	Upper-middle	South-East Asia	Thailand	6	0.2
24	Low-middle	South-East Asia	Bangladesh	5	0.2
25	Upper-middle	Europe	Russia	5	0.2
26	High	Eastern Mediterranean	Saudi Arabia	5	0.2
27	High	Europe	Switzerland	5	0.2
28	Low	Africa	Ghana	4	0.2
29	High	Eastern Mediterranean	Israel	4	0.2
30	High	Americas	Venezuela	4	0.2
31	Low-middle	Western Pacific	Vietnam	4	0.2
32	Upper-middle	Americas	Argentina	3	0.1
33	High	Europe	Belgium	3	0.1
34	Upper-middle	Western Pacific	China	3	0.1
35	Upper-middle	Americas	Colombia	3	0.1
36	Low-middle	Eastern Mediterranean	Egypt	3	0.1
37	High	Europe	Finland	3	0.1
38	High	Europe	Italy	3	0.1
39	High	South-East Asia	Maldives	3	0.1
40	High	Europe	Norway	3	0.1
41	High	Eastern Mediterranean	United Arab Emirates	3	0.1
42	High	Europe	Austria	2	0.1
43	Upper-middle	Americas	Brazil	2	0.1
44	High	Europe	Czechia	2	0.1
45	High	Europe	Denmark	2	0.1
46	Upper-middle	Americas	Dominican Republic	2	0.1

47	High	Europe	Greece	2	0.1
48	High	Europe	Latvia	2	0.1
49	Low	South-East Asia	Nepal	2	0.1
50	Upper-middle	Americas	Peru	2	0.1
51	High	Europe	Portugal	2	0.1
52	High	Eastern Mediterranean	Qatar	2	0.1
53	High	Europe	Sweden	2	0.1
54	Low-middle	Africa	Tanzania	2	0.1
55	High	Americas	Trinidad & Tobago	2	0.1
56	Upper-middle	Europe	Turkey	2	0.1
57	Low-middle	Europe	Ukraine	2	0.1
58	Low-middle	Africa	Zambia	2	0.1
59	High	Europe	Andorra	1	0.0
60	High	Americas	Anguilla	1	0.0
61	Upper-middle	Europe	Belarus	1	0.0
62	Low-middle	Western Pacific	Cambodia	1	0.0
63	High	Americas	Caribbean Netherlands	1	0.0
64	High	Americas	Chile	1	0.0
65	Upper-middle	Americas	Costa Rica	1	0.0
66	High	Europe	Cyprus	1	0.0
67	Upper-middle	Americas	Ecuador	1	0.0
68	Low-middle	Americas	Guyana	1	0.0
69	Low	Americas	Haiti	1	0.0
70	High	Western Pacific	Hong Kong SAR China	1	0.0
71	Upper-middle	Eastern Mediterranean	Iraq	1	0.0
72	High	Americas	Jamaica	1	0.0
73	High	Eastern Mediterranean	Kuwait	1	0.0
74	Upper-middle	Western Pacific	Mongolia	1	0.0
75	Upper-middle	Europe	Montenegro	1	0.0
76	Low-middle	Eastern Mediterranean	Morocco	1	0.0
77	Low-middle	Europe	North Macedonia	1	0.0
78	High	Eastern Mediterranean	Oman	1	0.0
79	High	Europe	Poland	1	0.0
80	Upper-middle	Europe	Romania	1	0.0
81	Low	Africa	Senegal	1	0.0
82	High	Western Pacific	Singapore	1	0.0
83	High	Americas	U.S. Virgin Islands	1	0.0
84	Low	Africa	Zimbabwe	1	0.0
				2440	100.0

The World Bank categorizes countries according to their gross national income per capita:

- Low-income: $1,135 or less.
- Lower-middle income: $1,136 to $4,465.
- Upper-middle income: $4,466 to $13,845.
- High-income: $13,846 or more.

Appendix 14. Diphtheria-related posts and cases by country and country classification

#	Diphtheria incid.	WHO region	Country	N. of posts	N. of cases
1	High	South-East Asia	India	67	49,739
2	High	Western Pacific	Philippines	29	710
3	High	Africa	Nigeria	18	4,159
4	High	South-East Asia	Indonesia	12	6,248
5	High	Eastern Mediterranean	Pakistan	17	2,641
6	High	Eastern Mediterranean	Yemen	7	4,681
7	High	South-East Asia	Thailand	6	389
8	High	Americas	Venezuela	4	1,761
9	High	South-East Asia	Nepal	2	2,541
10	High	Americas	Haiti	1	408
11	Low	Americas	United States	63	5
12	Low	Europe	United Kingdom	38	135
13	Low	Western Pacific	Australia	16	85
14	Low	Americas	Canada	10	31
15	Low	Europe	Ireland	6	2
16	Low	Western Pacific	Malaysia	4	133
17	Low	Africa	South Africa	3	32
18	Low	Europe	Germany	2	113
19	Low	Europe	France	2	110
20	Low	Europe	Ukraine	2	23
21	Low	Europe	Spain	2	8
22	Low	Eastern Mediterranean	Qatar	2	2
23	Low	South-East Asia	Bangladesh	1	110
24	Low	Western Pacific	New Zealand	1	9
25	Low	Europe	Czech Republic	1	5
26	Low	Western Pacific	China	1	2
27	Low	Western Pacific	Cambodia	1	0
28	Low	Europe	Israel	1	0
29	Low	Africa	Kenya	1	0
30	Low	Eastern Mediterranean	Kuwait	1	0
31	Low	South-East Asia	Maldives	1	0
32	Low	Americas	Mexico	1	0
33	Low	Europe	Montenegro	1	0
34	Low	Eastern Mediterranean	Oman	1	0
35	Low	Americas	Trinidad and Tobago	1	0