MODELING PRECIPITATION, ACUTE GASTROINTESTINAL ILLNESS, AND ENVIRONMENTAL FACTORS IN NORTH CAROLINA, USA

Kristen D. Downs

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Environmental Sciences and Engineering in the Gillings School of Global Public Health

Chapel Hill
2022

Approved by:

Dale Whittington

Gregory Characklis

Charles Konrad

Karen Levy

David Richardson

Anna Waller

Jason West

ABSTRACT

Kristen D. Downs: Modeling Precipitation, Acute Gastrointestinal Illness, and Environmental Factors in North Carolina, USA
(Under the direction of Dale Whittington)

Increasing intensity and frequency of extreme weather events due to climate change underscores the importance of understanding the influence of hydroclimatic variability on health. Meteorological drivers affect rates of acute gastrointestinal illness (AGI), but the association between precipitation and AGI, the sensitivity to modeling decisions, and the effects of sociodemographic and environmental risk factors are not well characterized. Furthermore, methodological differences may reduce inter-study comparability and can affect model estimates.

In this dissertation, we reviewed the methodologies of recent time series AGI-weather studies, including outcome and exposure variables, data sources, spatiotemporal aggregation, and model specification. To investigate the sensitivity of the association between AGI and precipitation to exposure definitions and effect measure modification (EMM), we used AGI emergency department (ED) visit and weather data (2008-2015) from North Carolina (NC) to develop daily, ZIP code-level quasi-Poisson generalized linear models and distributed lag models. We compared multiple precipitation metrics: absolute (total precipitation), extreme (90th, 95th, and 99th percentiles with and without zero-precipitation days), and antecedent (cumulative wet-dry days; 8-week wet-dry periods). We assessed for potential EMM by physiographic region, the density of hogs in concentrated animal feeding operations (CAFOs), and percent of population on private drinking water wells.

Depending on exposure definition, we observed an overall cumulative decrease of 1-18% in AGI ED rates following extreme precipitation events (over 0-7 days), with stronger effects associated with heavier rainfall, and a 2% (95% CI: 1.02, 1.03) increase after antecedent (8-week) wet periods. Inverse statewide results following extreme precipitation—dominated by the demographic weight of urban centers in the Piedmont region—were consistent with dilution effects posited by the concentration-dilution hypothesis but obscured dramatic sub-state variation. While EMM by private wells was inconclusive, region and hog density strongly modified the associations observed, with increased AGI ED rates following 95th percentile precipitation in the mountains (18%), coastal plains (19%), and areas exposed to hog CAFOs (7-15%).

Our results reveal the vulnerability of mountainous, coastal, and CAFO-impacted areas in NC to rainfall-exacerbated AGI risk. This dissertation highlights the hazards of data aggregation and importance of precipitation exposure definitions and effect measure modification when modeling climate-health relationships.

ACKNOWLEDGEMENTS

The work presented here was funded in part by the University of North Carolina Dissertation Completion Fellowship, the Carolinas Integrated Sciences and Assessment, and the Jack Kent Cooke Foundation Graduate Fellowship. I am grateful for the patient support of my advisor, Dale Whittington, my PhD committee, and the department of Environmental Sciences and Engineering (ENVR); of many fellow students throughout the years; and of my friends and family.

I dedicate this work to my parents, and particularly to my mom, Denise Dulion, for her resilience, unflagging support, and love, in addition to Wayne Garland and Lauren and the Fergusons. I have had the pleasure and privilege of being part of many communities in during my graduate work. These include many current and former students in ENVR, the Whittington research group, enduring friends from DoGEE, my Grenoblings, the Stuck to Unstoppable and Accountabilibuddies groups, Grace Community Church and All Saints Church. Thank you to AR El-Khattabi, Arbor Quist, Jess Lawson, Heather Taylor, Jodie Martin, and Elizabeth Christenson for your extra support during the push to finish this dissertation during the COVID-19 pandemic. There are many who have remained unnamed, but please know you're on my mind and in my heart. Your presence and participation in my life has brought me joy, taught me many lessons of the heart and mind, and lifted me through the difficult times.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

ACS	American Community Survey
AH	absolute humidity
AGI	acute gastrointestinal illness
AP	atmospheric pressure
AR	auto-regressive
ARIMA(X)	autoregressive integrated moving average (with exogenous variables)
ARMA	autoregressive moving average
AT	apparent temperature
BC	bagged CART
BSTHM	Bayesian space-time hierarchical model
CAFO	concentrated animal feeding operation
CART	classification and regression tree
CDF	cumulative distribution function
CI	confidence interval
CMA	China Meteorological Administration
CONUS	continental United States
CSO	combined sewer overflow
DA	data assimilation
DALY	disability-adjusted life year
DEM	digital elevation model
DLM	distributed lag model
DLNM	distributed lag non-linear model

DM	data mining
DOW	day-of-week
DTR	diurnal temperature range
DW	drinking water
ECMWF	European Centre for Medium-Range Weather Forecasts
ED	emergency department
EMM	effect measure modification
ENSO	El Niño - Southern Oscillation
EPA	US Environmental Protection Agency
EPT	extreme precipitation (percentile)
ESRI	Environmental Systems Research, Inc.
ETT	extreme temperature (percentile)
GA	Georgia
GAM	generalized additive model
GAMM	generalized additive mixed model
GCD	gridded climate data
GDP	gross domestic product
GEE	generalized estimating equation
GI	gastroenteritis
GIS	geographic information system
GLM	generalized linear model
GLMM	generalized linear mixed model
GWD	gridded weather data
HCGI	highly credible gastrointestinal symptoms

HI	heat index
HU	hydrologic unit
IAM	integrated assessment model
ICD-10-CM	International Classification of Diseases–Clinical Modification, 10th revision
ICD-9-CM	International Classification of Diseases–Clinical Modification, 9th revision
IDP	internally displaced persons
IID	intestinal infectious disease
IRR	incidence rate ratio
LDAS	land data assimilation system
MA	moving average
MARS	multivariable adaptive regression splines
MAUP	modifiable areal unit problem
MLE	maximum likelihood estimation
MTUP	modifiable temporal unit problem
NB	negative binomial
NBRM	negative binomial regression model
NC	North Carolina
NC DETECT	North Carolina Disease Event Tracking and Epidemiologic Collection Tool
NC DPH	North Carolina Division of Public Health
NDVI	normalized difference vegetation index
NJ	New Jersey
NOAA	U.S. National Oceanic and Atmospheric Administration
NPDWR	National Primary Drinking Water Regulations
NYSE	New York Stock Exchange

OLS	ordinary least squares
OTC	over-the-counter
PHEP	Public Health Emergency Preparedness
PIM	population intervention model
PPT	precipitation
PRISM	Parameter-elevation Regressions on Independent Slopes Model
QMRA	quantitative microbial risk assessment
RF	random forest
RH	relative humidity
RR	rate ratio
RTA	regression tree analysis
SARIMA(X)	seasonal autoregressive integrated moving average (with exogeneous factors)
SBI	symmetric bidirectional
SES	socioeconomic status
SH	specific humidity
SIWR	susceptible-infectious-water-recovered
SSO	sanitary sewer overflow
SST	sea surface temperature
Tmax	maximum temperature
Tmean	mean temperature
Tmin	minimum temperature
TS	time series
TSR	time series regression

UNC CCHI	University of North Carolina at Chapel Hill Department of Emergency Medicine's Carolina Center for Health Informatics
UOA	unit-of-analysis
US, USA	United States of America
VAR	vector autoregression
VP	vapor pressure
WASH	water, sanitation and hygiene
WCRP	World Climate Research Program
WHO	World Health Organization
WMO	World Meteorological Organization
WS	wind speed
ZCTA	ZIP code tabulation area
ZIP code	Zone Improvement Plan code (postal code used by the US Postal Service)

1. INTRODUCTION AND BACKGROUND

1.1. Motivation

1.1.1. General context

Global climate change is expected to have profound impacts on environmental and hydroclimatic systems worldwide, with important ramifications for human health (Patz et al., 2014). In addition to increased average surface temperatures, impacts will include changes in the intensity, frequency, and duration of precipitation and extreme events such as heat waves, droughts, storms, storm surges, and floods (IPCC, 2013). Among the indirect effects of climate change, shifts in microbial transmission and contamination are projected to exacerbate morbidity and mortality due to waterborne gastrointestinal diseases, which are sensitive to temperature and hydroclimatic conditions (J. N. S. Eisenberg, Desai, et al., 2007).

The transmission of enteric pathogens that cause gastrointestinal illness can be affected by temperature, humidity, precipitation, runoff, flooding, drought, and storms (K. Levy et al., 2016). Higher temperatures can increase the replication rate of bacterial pathogens, while heavy precipitation and flooding can carry microbial contamination into water supplies, creating challenges for engineered systems that are intended to disrupt disease transmission pathways or manage water resources (K. Levy et al., 2016). In the U.S., approximately 7.15 million cases of waterborne infectious diseases are estimated to occur each year, resulting in $3.33 billion in direct healthcare costs (Collier et al., 2021).

The WHO (2014) projects that rising temperatures alone will cause 48,000 additional diarrheal deaths per year globally between 2030-2050. No corresponding estimate is available for how changing precipitation might affect global disease burden: uncertainties in the association between gastrointestinal illness and precipitation, compounded by uncertainties about the effects of climate change on local precipitation, have so far prevented any such projections (Hales et al., 2014). However, it is important to understand how changes to hydroclimatic drivers may affect disease risk as small relative changes have the potential for correspondingly large overall increases in diarrheal disease burden (K. Levy et al., 2016).

1.1.2. Acute gastrointestinal illness and diarrheal diseases

Enteric pathogens are those that cause gastrointestinal distress when ingested. These organisms are transmitted from person to person primarily through the fecal-oral route (Bylund et al., 2017; Julian, 2016; Kotloff, Nataro, Blackwelder, & Nasrin, 2013; Lanata et al., 2013; Prüss-Ustün et al., 2014; Roy et al., 2006; Scallan, Hoekstra, et al., 2011; Wolf et al., 2014). Exposure can occur through interdependent pathways mediated by the environment and contaminated water, food, hands or hygiene practices, surfaces (fomites), and insects carrying pathogens (J. N. S. Eisenberg, Scott, et al., 2007).

Acute gastrointestinal illness (AGI) or acute gastroenteritis describes a set of illnesses of the intestine caused by multiple enteric pathogens and is loosely defined by symptoms of diarrhea, vomiting, and nausea (Roy et al., 2006), but there is no consistent terminology in the literature: a high degree of overlap exists between terms like AGI, intestinal infectious disease (IID) (Garthright et al., 1988; Roderick et al., 1995), highly credible gastrointestinal symptoms (HCGI) (Payment et al., 1991), enteric illness, diarrheal disease, waterborne (Leclerc et al., 2002) and foodborne diseases (Scallan, Griffin, et al., 2011; Scallan, Hoekstra, et al., 2011), and the like. Vocabulary typically reflects whether researchers reference the exposure route, the primary symptoms, or the location of the infection in the body. Without a universal medical definition, measurement of these infections is likewise inconsistent. Diarrhea (frequency and severity) is the most frequently studied symptom of

AGI (Roy et al., 2006), but even this measure is without standardization. The WHO (World Health Organization (WHO), 2017) defines diarrhea "as the passage of three or more loose or liquid stools per day (or more frequent passage than is normal for the individual)," but many studies use different definitions. Studies frequently include consideration of nausea, vomiting, abdominal pain or cramps, and/or systemic symptoms related to intestinal infection when looking at AGI, and in this way, AGI is sometimes seen as a more general term that encompasses diarrheal disease.

In this dissertation, we are not interested in the nuance of the distinctions mentioned above; terms are used interchangeably unless otherwise specified. We are interested in the entire set of enteric diseases resulting in a variety of symptoms and caused by a number of infectious pathogens, including enteric viruses (e.g., norovirus, rotavirus), bacteria (e.g., *E. coli*,[1] *Shigella* spp., *Campylobacter spp., Salmonella spp.*), and protozoa (e.g., *Cryptosporidium* spp., *Entamoeba histolytica*, *Giardia lamblia*) (Leclerc et al., 2002; Scallan, Hoekstra, et al., 2011).

Although morbidity and mortality due to diarrheal diseases have decreased over the last few decades, they remain a significant cause of illness and death globally. In 2015, diarrheal diseases resulted in an estimated 1.31 million deaths, 71.59 million disability-adjusted life years (DALYs) and 23.923 billion episodes[2] among all ages worldwide (Troeger et al., 2017). Globally, the burden of diarrheal disease falls most heavily on children under 5, with about half a million deaths estimated in 2015, and on populations in low- and middle-income countries (Troeger et al., 2017).

[1] There are six strains of pathogenic enteric (a.k.a. diarrheagenic) *Escherichia coli*: EPEC (enteropathogenic) enteropathogenic *E. coli* (EPEC), Shiga toxin-producing *E. coli* (STEC) (e.g., enterohemorrhagic *E. coli* [EHEC]), *Shigella*/enteroinvasive *E. coli* (EIEC), enteroaggregative *E. coli* (EAEC), diffusely adherent *E. coli* (DAEC), and enterotoxigenic *E. coli* (ETEC), and adherent invasive *E. coli* (AIEC) (Croxen et al., 2013; Kaper et al., 2004; Nataro & Kaper, 1998). EPEC and ETEC infections are associated with significant child mortality in lower- and middle-income countries (LMICS) (Kotloff, Nataro, Blackwelder, Nasrin, et al., 2013).

[2] 95% uncertainty intervals reported at 1.23-1.39 million deaths, 66.44-77.21 million DALYs, and 23.01-25.03 billion episodes. The DALY is defined as "[t]he sum of years of potential life lost due to premature mortality and the years of productive life lost due to disability." (http://www.who.int/mental_health/management/depression/daly/en/)

High-income countries also experience problems with drinking water safety and a health burden from acute gastrointestinal illness, particularly norovirus, (Beaudeau, Schwartz, et al., 2014; Flahault & Hanslik, 2010; Kowalzik et al., 2015; Lake et al., 2007; Lopman et al., 2003; Naumova et al., 2005a; Tinker et al., 2009; Zmirou et al., 1995). The economic impacts of these illnesses include direct healthcare costs and losses in employee productivity (Hutton & Haller, 1994). Recent analyses for the United States have estimated an annual gastrointestinal illness burden ranging from 4.3 to 16.4 million cases per year due to contaminated drinking water (Colford et al., 2006; Messner et al., 2006).

1.1.3. Complexity of diarrheal disease mechanisms and causal pathways

Like many other infectious diseases, those with diarrheal symptoms often have complex mechanisms and causal pathways. Seasonal, climatic, and bio-physical factors interact to create system dynamics that are difficult to disentangle (Altizer et al., 2006; Fisman, 2007; K. Levy et al., 2016; Lo Iacono et al., 2017; Mellor et al., 2016).

Seasonal drivers of infectious diseases are many, and include host population behaviors, pathogen-pathogen interactions, and numerous environmental factors (Altizer et al., 2006; Fisman, 2007). Seasonally-variable environmental conditions affect the abundance, survival, and virulence of pathogens; host susceptibility and immune defense; timing of reproduction/availability of susceptible hosts; and spatial patterns of disease (Altizer et al., 2006; Fisman, 2007). These factors are not independent; they co-occur (Fisman, 2007).

The multiple, interdependent pathways that underlie diarrheal diseases (J. N. S. Eisenberg et al., 2012) further complicate the relationship between infectious diseases and environmental conditions, whether the latter reflect seasonality (Altizer et al., 2006; Fisman, 2007) or weather and climate (K. Levy et al., 2016; Lo Iacono et al., 2017; Mellor et al., 2016). For example, viral gastroenteritis is generally thought to occur during winter or cooler months, while bacterial and protozoan gastroenteritis occurs in summer or warmer months. Yet each diarrheagenic species displays distinct and more complex seasonality patterns, which also vary geographically.[3]

Although this dissertation focuses on the short-term impacts from weather or meteorological factors instead of the impacts of long-term climate or climate change, understanding meteorological factors based on past observations is key to understanding the future effects of climate change. In the climate-health literature, terms including weather, meteorological, climatic, climatological, hydrometeorological or hydroclimatological[4] factors are often used interchangeably to refer to the impacts of temperature and precipitation (most common), in addition to other variables (e.g., relative or specific humidity, dewpoint temperature, apparent temperature, sea surface temperature (often used as a predictor for cholera), etc.). Transmission routes for gastrointestinal illnesses, particularly those related to water quality or quantity, are affected by meteorological and hydroclimatological variables including temperature, precipitation, runoff, flooding, drought, humidity, and storms (Bylund et al., 2017; J. N. S. Eisenberg, Scott, et al., 2007; K. Levy et al., 2016; Mellor et al., 2016). Water scarcity limits access to a clean, safe source of drinking water or adequate amounts of water for personal hygiene (Stelmach & Clasen, 2015). Conversely, too much water can increase the availability of unsafe drinking water sources (Hunter & Wang, 2010), contaminate water supply systems (Carlton et al., 2014; Febriani et al., 2010; Gleason & Fagliano, 2017; Teschke et al., 2010; Tornevi et al., 2016; Uejio et al., 2014), or bring exposure to contaminated floodwaters (Watson et al., 2006).

[3] Multiple studies have investigated the seasonality of enteric pathogens, including rotavirus (e.g., Cook et al., 1990; Cunliffe et al., 1998; Haffejee, 1995; Jagai, Sarkar, et al., 2012), norovirus (e.g., Ahmed et al., 2013; Lopman et al., 2009, 2011; Rohayem, 2009), diarrheagenic *E. coli* (e.g., Philipsborn et al., 2016), *Cryptosporidium* (e.g., Jagai, Sarkar, et al., 2012) and other pathogens leading to diarrhea and AGI (Chao et al., 2019; Chui et al., 2011; Desai et al., 2012; Drayna et al., 2010; Jagai, Griffiths, et al., 2012; Leclerc et al., 2002; Naumova et al., 2007).

[4] Hydroclimatology is defined as the "study of the influence of climate upon the waters of the land" (Langbein, 1967) and includes "hydrometeorology as well as the surface and near surface water processes of evaporation, runoff, groundwater recharge, and interception" (Wendland, 1987).

Changes in water quality mean changes in the aqueous concentrations of constituents and contaminants of interest (e.g., pathogens). Concentration reflects both the mass (or count) of contaminants present, and the volume of water. Precipitation-related factors affect both the numerator and the denominator. There are many hydroclimatically-mediated transmission pathways that can impact diarrheal disease risk (Bylund et al., 2017; K. Levy et al., 2016). For example, rainfall events of varying intensity, frequency, and duration can mobilize contaminants from point- and non-point sources (Kraay et al., 2020); sanitary and combined sewer overflows following heavy rain events release fecal matter (Jagai et al., 2015, 2017); and flooding can further mobilize and spread microbial contaminants (Carroll et al., 2010; de Man et al., 2016; Quist, Fliss, et al., 2022; Reacher et al., 2004; Schnitzler et al., 2007; Wade et al., 2014) leading to increased cases of AGI.

These environmental processes can enhance the transfer of microbial contaminants into water supplies. At the same time, they affect the volume of water available. Taken together, weather can affect the levels of microbial contaminants in water in three different ways: the so called "runoff effect," the dilution effect, and the concentration effect (Kraay et al., 2020; K. Levy, Hubbard, Nelson, et al., 2009). As summarized in Levy *et al.* (2009), the runoff effect results in increased microbial contamination when wet conditions flush fecal material into water bodies and/or mobilize bacteria that reside in soil. It operates at both seasonal and short-term (e.g., hourly) scales. The dilution effect describes how a large volume of clean meteoric water from a pulse of rainfall will dilute contaminant concentrations as it moves through streams and water bodies over the course of days or weeks. The concentration effect occurs at seasonal scale, during the dry season (or during droughts) when contaminant concentrations increase because the total volume of water available falls. Regardless, microbial contaminants mobilized, diluted, or concentrated may come from zoonotic sources, poor sanitation practices, survival and/or growth in soils, and from use of water bodies for activities like bathing and washing. Kraay and colleagues (2020) provide an in-depth systematic review that evaluates and supports the evidence for the concentration-dilution effect based on precipitation and diarrheal disease studies.

Further research is needed to understand relevant associations between diarrheal disease and weather, as well as potential modifiers, confounding factors, and corresponding uncertainties. A stronger understanding of the relationship between diarrheal diseases and antecedent weather conditions can be used to improve modeling and inform future intervention efforts. Ultimately, understanding and projecting climate and disease dynamics to help protect human health will require research advances within the three major categories of climate-health studies (McMichael & Lindgren, 2011): empirical climate-disease studies on the impacts of recent variations and trends in climatic variables (traditional epidemiological research); empirical observations of actual changes in known or plausible climate-sensitive health outcomes and statistical studies estimating the attributable burden of disease; and climate change scenario-based modeling estimating future health risks in specified regions or populations. This dissertation contributes to the second research areas through investigation of empirical relationships between weather and acute gastrointestinal illness.

1.2. Linking climate change with diarrheal disease

In a meta-analysis of temperature and diarrheal disease studies, Carlton and colleagues (2016) found a positive association between ambient temperature and all-case diarrhea with a pooled incidence rate ratio (IRR) of 1.07 (95% confidence interval (CI) 1.03-1.10). The positive association held for bacterial diarrhea (IRR 1.07; 95% CI 1.04-1.10), but the association for viral diarrhea varied across studies (0.96; 0.82-1.11). Chua and colleagues expanded the former with a meta-analysis of ambient temperature and enteric pathogen-specific illnesses studies published between 2000-2019 that are consistent with the trends in earlier studies (positive associations for bacterial and inverse for viral enteritis). They reported overall relative risks of incidence per 1°C increase in temperature for bacterial—salmonellosis (1.05; 1.04-1.07), shigellosis (1.05; 1.04-1.10), campylobacteriosis (1.02;

1.01-1.04), cholera (1.04; 1.01-1.07), *Escherichia coli* enteritis (1.04; 1.01-1.07), and typhoid (1.15; 1.07-1.14)—and viral—rotaviral (0.96; 0.90-1.02), noroviral (0.89; 0.81-0.99)—enteritis.

Building on such associations, studies have included temperature when projecting the global impacts of climate change on diarrheal disease. Modelers have used both mean temperature change (Anthoff & Tol, 2013b; Kolstad & Johansson, 2011) and average annual temperature anomaly (Hales et al., 2014) to represent the impacts of climate change on their projections of gastrointestinal illness. Results are often uncertain. For example, the uncertainty for predictions of regional changes in relative risk of diarrhea between 2010-2099 simulated over a range of climate change model temperature scenarios was driven not by the climate model selected, but by the uncertainty in the relationship[5] between diarrhea and temperature (Kolstad & Johansson, 2011).

Precipitation exposures vary more than temperature, and there are many options for quantifying precipitation, including total rainfall, percentile indices for "extreme" rainfall, and multiple metrics—developed to study concentration-dilution effects, see (Kraay et al., 2020)—that consider heavy rainfall events following wet or dry periods. While many AGI-weather studies include few covariates or none at all, some recent studies have explored effect modification of AGI and weather by multiple age groups (e.g,, J. Cheng et al., 2017; Chou et al., 2010; Gleason & Fagliano, 2017; Wangdi & Clements, 2017), water sources (de Roos et al., 2020; Gleason & Fagliano, 2017; Teschke et al., 2010; Uejio et al., 2014), farming activities (Febriani et al., 2010; Quist, Holcomb, et al., 2022), region (Jiang et al., 2015; D. Lee et al., 2019; Soneja, Jiang, Upperman, et al., 2016), season (Gleason & Fagliano, 2017; Kraay et al., 2020), and El Niño-Southern Oscillation phases (Lama et al., 2004). In short, the relationship between gastrointestinal illness and precipitation is not well characterized as temperature, nor are the risk factors that may confound or mediate the

[5] In the simulation, this relationship was represented by a temperature-diarrhea coefficient alpha (α), defined "as the estimated percent increase in the relative risk (RR) of diarrhea with each 1°C temperature increase."

effects of changing climate and weather on the incidence of all-cause or pathogen-specific diarrheal illnesses (Kraay et al., 2020).

Empirical studies of the effects of historic variations and trends in climatic variables on health in different locations are foundational to understanding the complex relationships and risk disparities, and improving climate change impact projections, adaptation, and resilience. However, many epidemiological studies vary by setting, data source, methodology, and statistical or analytical technique, making it difficult to determine whether differences in published results reflect modeling decisions or characteristics of the local setting. In all cases, a dependence on past data for future modeling assumes stationarity of the relationships involved.

Any projections of morbidity and mortality due to acute gastrointestinal and diarrheal diseases must account for opposing drivers. Factors such as economic development, better healthcare services, and improvements in water and sanitation infrastructure can decrease the incidence and burden of diarrheal disease. These factors are dynamic and heterogeneous in both time and space (Fuente et al., 2020; Jeuland et al., 2013). Conversely, increases in adverse weather conditions and extreme events due to climate change may raise the rate of gastrointestinal and diarrheal disease (Anthoff & Tol, 2013b; Hales et al., 2014; Kolstad & Johansson, 2011b).

Given that these opposing drivers are in flux, it is important to understand their relative magnitude and the interactions between them: how does weather affect diarrheal disease, and how might this relationship change under evolving economic, infrastructural, and demographic conditions? Consideration of spatiotemporal scale is particularly important because of the seasonality of infectious disease, and since weather and climate impacts occur locally and regionally, whereas spatial averaging or aggregation reduces variability and can systematically bias the measure of climatic factors.[6] Improving the reliability of future diarrheal disease projections in the face of a changing climate will depend on an improved understanding of the associations between diarrheal diseases and climatic factors like temperature and precipitation, and on robust modeling.

Relationships between diarrheal diseases and antecedent temperature and rainfall

The relative risk of all-cause diarrheal disease is estimated to increase by 3-11% for every 1ºC increase in temperature (Carlton et al., 2016; Checkley et al., 2000; Chou et al., 2010; Hashizume et al., 2007; Lama et al., 2004; Onozuka et al., 2010; Singh et al., 2001; Y. Zhang et al., 2008b). However, when studies are disaggregated by the pathogen etiology, a more complex story emerges. Temperature is generally positively associated with bacterial and protozoan acute gastrointestinal illness (AGI), negatively associated with viral AGI, and positively associated (with a smaller effect size) with all-cause AGI (K. Levy et al., 2016).

The relationship between precipitation and diarrheal disease is less clear, but still important. There is more variation amongst rainfall studies than temperature studies, with increases (most common), decreases, and non-significant changes in relative risk with increased rainfall depending on the local context and study (Guzman Herrador et al., 2015; K. Levy et al., 2016). As the runoff, dilution, and concentration effect theories suggest (Kraay et al., 2020; K. Levy et al., 2016), the complexity and variability in results in rainfall studies may be related to the intensity, frequency, and duration of rainfall in a given hydroclimatological context and watershed system. Between 1948 and 1994, 51% and 68% of U.S. waterborne disease outbreaks occurred following precipitation events above the 90th and 80th percentiles respectively (Curriero et al., 2001).

A limited number of more recent studies have attempted to investigate how antecedent rainfall patterns, particularly heavy rainfall, have differing effects on disease outcomes. For example,

[6] Aggregation bias has been studied in multiple fields, including in the context of climate change economics and policy with respect to the integrated assessment models (Schumacher, 2018) and agriculture (Fezzi & Bateman, 2015).

a recent New Jersey study by Gleason and Fagliano (2017) investigated the effect modification of season and drinking water source on the association between 90th percentile precipitation and hospitalizations for AGI, finding positive associations during the warm season for surface water systems and 'other' water source systems (e.g., small community water systems, private wells, unknown), but not public groundwater systems. Conversely, during the cold season, AGI and heavy rainfall were inversely associated in groundwater and surface water systems (Gleason & Fagliano, 2017d). In two epidemiological studies in Ecuador, Bhavnani *et al.* (2014) found the odds of diarrhea to be highest after maximum and minimum rainfall respectively, depending on water and sanitation conditions (unimproved water source and unimproved sanitation), and Carlton *et al.* (2014) found heavy rainfall events to be associated with increased diarrheal incidence following dry periods, as well as wet periods.

Currently, there is no clear *mechanistic* description of how climatic factors (temperature, precipitation) and extreme events affect diarrheal diseases (Mellor et al., 2016). The magnitude of the uncertainty around relative risk estimates involving temperature and rainfall is also unclear. Local watersheds, hydroclimatological conditions, and infrastructure are important; the influence (in magnitude and direction) of precipitation and temperature levels and changes are likely to vary by location. Studies are also difficult to compare because of the wider variation in methodologies, study conditions and geography, and statistical methods (Bylund et al., 2017; K. Levy et al., 2016).

1.3. Modeling climate-disease studies in environmental epidemiology

1.3.1. Study variability in modeling approaches and specification

Modeling is an essential tool in both epidemiology and climate research. But many epidemiological studies use different methodologies, data sources, statistical techniques, variable selection, and spatial and temporal resolution, so it can be difficult to compare analytical models and their results. Without attention to these details, it is impossible to say whether varied outcomes reported by different researchers reflect real differences or result from differences in study design (Butler & Hall, 2009). This section summarizes some of the methodological differences in employed in epidemiological studies of weather-diarrheal disease. A literature review that expands on these subjects in more detail and with specific examples is available in Chapter 0.

Study types and statistical models

Most of the common epidemiological study types are represented in the literature on the relationship between diarrheal disease and environmental factors. These include observational studies (which include ecological studies), case-control studies, case-crossover studies, cross-sectional studies, cohort or panel studies, and experimental studies such as randomized control trials and household interventions (Munnangi & Boktor, 2018). Bylund *et al.* (2017) systematically reviewed recent research on sporadic gastroenteritis associated with drinking water in high-income Northern hemisphere settings and described the advantages and disadvantages of most epidemiological study designs. The authors reported that it was too challenging to conduct a meta-analysis, because many of the studies reviewed lacked statistical robustness, and because local variations made comparisons difficult.

Observational study designs are also common in weather-health studies. Model choice is typically a function of study design and data, including various types of time series models and, less frequently, case-crossover models. Time series analysis is useful in observational studies as it controls for time-invariant factors, but its application to infectious diseases adds complications, including nonlinearities, that often need to be addressed by more complex model specification. Imai *et al.* (2015) identified five common issues in time series modeling for infectious diseases and weather, and recommended potential ways to approach each: (1) changes to immune populations; (2) strong autocorrelations; (3) wide range of plausible lag structures and associated patterns; (4) seasonal and long-term trend adjustments; and (5) large overdispersion. Most of the approaches they mention are represented in AGI-weather studies, along with more common time series models such as Poisson generalized linear models (GLM) and generalized additive models (GAM).

The variety of statistical model types and specifications in weather-health studies likely represents a source of variation that makes it more difficult to compare results between studies.

Study design

The design of a weather-health study is influenced by the availability of outcome and exposure data sources, the study population of interest, and the unit(s) of analysis. Differences in these three factors across studies contribute to inter-study variation.

Health data used in epidemiological studies is commonly sourced from pharmacies, telephone triage, health care facilities, questionnaires, health diaries, and patient registers. The source of health data often affects the study methodology or vice versa, whether including ecological, case-control, cross-sectional, cohort, and household interventions studies (see TABLE 1 for data sources matched to study methodology based on Bylund et al. (2017)). Each data source has advantages and limitations in terms of availability, risk of bias, risk of under-reporting, and information available on health outcomes, demographics, and potential risk factors (for details refer to Bylund et al., 2017).

Depending on data availability and statistical power, studies must decide whether to focus on a segment of the population by age or to disaggregate their data. Age is an important demographic factor: children, especially those under 5, are more vulnerable to and have a higher prevalence of diarrheal diseases (Fischer Walker, Aryee, et al., 2012; Fischer Walker, Perin, et al., 2012; Pires et al., 2015).

The unit of analysis also varies from study to study, with implications for the type(s) of analysis that may be conducted, as well as the interpretation of results. As reviewed in Chapter 2 of this dissertation, most weather-diarrhea studies are conducted at the individual or household level. Other units of analysis may range from community to national, or may adopt a population-weighted grid scale.

TABLE 1. Data sources commonly used within study methodologies common to studies investigating the relationship between acute gastrointestinal illness (AGI) and drinking water or weather. *Source*: Bylund *et al.* (2017, n. Table 3, columns 1 & 3)

STUDY METHODOLOGY	DATA SOURCE: *Pharmacies*	*Telephone triage*	*Health care*	*Questionnaires*	*Interviews*	*Health diaries*	*Patient registers*
Ecological study	X	X	X				
Case-control study			X				
Cross-sectional study				X	X		
Cohort study				X	X	X	X
Household intervention				X	X	X	

Note: This table was developed based on data sources used in study methodologies common in diarrheal disease studies, as identified by Bylund et al. (2017, n. Table 3, columns 1 & 3).

Variable selection

Variable selection for outcomes, weather factors, and other risk factors or covariates is also an important modeling decision with implications on study results (validity, bias, generalizability, etc.) and variation between studies.

With respect to health outcome, AGI is complicated because it is a composite of enteric illnesses with multiple causative agents (viral, bacterial, and protozoan), each with different disease mechanisms, seasonal cycles, and relationships with weather. Scientists studying diarrheal diseases in human populations would ideally prefer laboratory-tested samples that confirm the etiology or etiologies of those exhibiting and not exhibiting symptoms. However, laboratory-confirmed data is often in low supply and can be prohibitively expensive in terms of time and labor. While some well-funded or healthcare facility studies are able to identify and, therefore, disaggregate diarrheal cases by their causative agent, many define their outcome as all-cause diarrhea due to lack of available data or use ICD-9-CM/ICD-10-CM codes. Unfortunately, aggregating cases together into all-cause diarrhea can lead to an attenuation in results if different pathogens have different relationships with a weather variable. This is evident in the case of temperature where viral AGI tend to be negatively correlated with temperature and AGI of bacterial and protozoan origin tend to be positively correlated with temperature, while all-cause AGI is often positively correlated with temperature to a lesser degree (Carlton et al., 2016; Chua et al., 2022; K. Levy et al., 2016).

The measurement and characterization of weather variables also introduces challenges. First, based on the study's hypotheses, local environment, and available weather data, researchers must select the weather variable(s) include in the model(s) and decide whether to include multiple exposure variables in the same model. Weather data may come from a range of sources, including weather stations, field collection with rain gauges, and modeled or reanalysis data. The second challenge is defining and measuring the exposure variable(s). Weather variables could be represented by the mean, minimum, maximum, percentile (e.g., 90^{th}, 95^{th}, or 99^{th} percentile rainfall), or by various composite measures such as degree days, a smoothed/interpolated spline, and with or without lags. Selection of a variable well-suited to the local context tends to improve the validity of results, but it makes comparing results across studies difficult, especially in the case of precipitation studies where there is more variation in the definitions of rainfall or heavy rainfall than in temperature studies (K. Levy et al., 2016). Finally, the choice to include other risk factors, whether as controls for confounding factors or effect modification, is important to minimize missing variable bias and more accurately model the relationship between dependent and independent variables (as appropriate for a given statistical model), but provides another source of study variation and quality. Covariates and risk factors may include physical variables describing the environment; information on water, sanitation, and/or hygiene infrastructure; or conditions, demographics, and socioeconomics. The inclusion of risk factors, while desirable, is often limited by data availability. For example, healthcare data often has dataset or privacy limitations that inhibit the inclusion of or sharing of case data that would provide context through relevant risk factors. Overall, it is evident that variable selection and measurement is a source of study variation, but the extent of its influence when comparing study results is unknown.

1.3.2. Spatio-temporal scale and resolution

How we model within space and time is important, in terms of both scale and resolution. When considering the spatial unit(s) of analysis, smaller units (i.e., more resolved) are generally better because it is desirable for the study population and unit(s) of analysis to be as close as possible to the scale at which the relationships of interest occur. It can be difficult to assess whether the correct unit(s) have been selected in a given study and, therefore, whether the results are accurate. Although this problem is only infrequently discussed in public health, epidemiology, and environmental fields, it is well known in the fields of geography and spatial analysis as the modifiable areal unit problem (MAUP), which was formally developed Openshaw and Taylor (Openshaw, 1978, 1984b; Openshaw & Taylor, 1979).

The MAUP is "a problem arising from the imposition of artificial units of spatial reporting on continuous geographical phenomena resulting in the generation of artificial spatial patterns" (Heywood et al., 1988). There are two types of MAUPs: scale (or aggregation) and zone (or grouping) (Fotheringham & Wong, 1991; Openshaw & Taylor, 1979; D. Wong, 2009; D. W. S. Wong, 2004). Issues of scale and zoning commonly arise in human geography and demography, notably from artificial boundaries such as census tracts, municipalities, or counties. Though less often directly addressed, MAUP is also problematic to physical geography and natural science research or applications involving remote sensing and geographic information systems (GIS) (Dark & Bram, 2007). The MAUP likewise applies in epidemiological and weather-health studies, which also involve a combination of human and natural system data, because one must consider whether and how much to aggregate or group the outcome and exposure by political boundaries or ecological "zones." When using secondary data, researchers may have little control over the original level of aggregation or method of zoning or grouping depending on the data source.

Tradeoffs between spatio-temporal resolution and sample sizes have led to two general approaches to estimating the relationship between weather and health. On one extreme are local epidemiological or economic time series studies that are conducted on small spatial scales (e.g., hospital, community, or city) over short time steps (e.g., monthly, daily, weekly) for diarrheal morbidity or infant/child mortality (e.g., as reviewed in Carlton et al., 2016; Chua et al., 2022; Kraay et al., 2020; K. Levy et al., 2016). These smaller studies may be able to access more resolved or detailed outcome or exposure data, but they have limited generalizability. On the other extreme are studies that look at the national or regional effects of weather or climate change at longer time scales, as in Integrated Assessment Models (IAMs) such as the global DICE or regional RICE model (Nordhaus, 2010; Nordhaus & Sztorc, 2013), the FUND model (Anthoff & Tol, 2013b, 2013a), and the PAGE model (Hope, 2011, 2013). Due to improved spatial coverage, these studies may have increased generalizability, but may suffer in terms of data detail, resolution, or frequency that can lead to aggregation bias.

Some studies (e.g., Bandyopadhyay et al., 2012; Kudamatsu et al., 2012). have used country-wide georeferenced Demographic Health and Surveys (DHS; https://dhsprogram.com), which are advantageous because they provide nationally representative surveys in lower- and middle-income countries in one or more years, and attempt to overcome the limitations of the cross-sectional surveys by weather data to the survey date or date of birth.

A few studies have simultaneously captured a larger sample size, greater representativeness, and refined spatial resolution using global gridded spatial data (Hales et al., 2014) or study areas in multiple counties aggregated to the annual level (Lloyd et al., 2007). However, while most epidemiological studies and meta-analyses of monthly temperature and diarrheal disease in small study areas have found positive associations with temperature, Lloyd *et al.* (2007) found a non-significant relationship between mean temperature and diarrheal rates (episodes per child-year). The authors used 36 studies with at least one year of two-week diarrheal disease surveillance data and aggregated the study data to the annual level. If not reflective of a true lack of relationship, one may hypothesize that the null results may be due to the annual temporal aggregation. Refer to Chapter 0 for further exploration of AGI-weather studies and their spatial and temporal coverage and resolution.

The variation in spatial and temporal scale and resolution of available outcome and exposure data and model development at a given spatio-temporal scale is a critical challenge to understanding and projecting the impacts of weather variability and climate change on diarrheal diseases. Limitations of available data and potential mismatches between the temporal and spatial resolution and representativeness of outcome and exposure data of may contribute to methodological differences between studies. For example, health data may be gathered at the hospital, city, regional, or national level, while weather data is available from a scattering of weather stations or in gridded form. Not all of this data is available at fine temporal resolution (high frequency: sub-daily, daily, or weekly) and/or for a long periods of record in all locations. Time is especially important for studies of infectious diseases like AGI, because these infections tend vary in likelihood and severity over time. Seasonality and interannual variability, therefore, are important to account for when modeling infectious diseases (Kelly-Hope & Thomson, 2008). High resolution data temporal are also important to capture short-term extreme weather like flooding, heavy rains, and heatwaves. These events can be important drivers for disease and are typically lost in temporal aggregation over longer periods of time.

In summary, the spatial and temporal scale—both unit size and expanse in space or length in time—of health outcome data, climatic data, and the final analysis are important to consider for the following reasons: (1) infectious diseases are seasonal (Fisman, 2007; Kelly-Hope & Thomson, 2008); (2) extremes are likely to impact health more than averages (Patz et al., 2000); (3) there may be nonlinear relationships between diseases and their risk factors (e.g., Gasparrini et al., 2017); (4) weather and climate impacts occur locally and regionally; and (5) spatial averaging reduces variability and systematically biases the measure of climatic factors (Brown et al., 2013; Hanemann et al., 2014). There are many reasons why spatio-temporal resolution is important, but the challenge remains of collecting and integrating data of sufficient resolution so as to reveal meaningful relationships between antecedent weather and infectious diseases, as well as comparing between analyses or analyzing data together to understand how study results may reflect local contextuality and broader generalizability.

1.4. North Carolina

The U.S. state of North Carolina (NC) is well-suited to research on associations between weather and AGI, thanks to its climatological, geographic, and sociodemographic profile. It has a large, diverse, and growing population, and exhibits significant variation in topography and weather. Statewide, about 29,000 emergency department visits and $40 million in associated costs are attributed to microbial contamination of drinking water annually (DeFelice et al., 2016). Agricultural non-point sources likely contribute to contamination issues. The availability of health and climate data from local, state (e.g., NC Department of Health and Human Services), NC Department of Environmental Quality, NCDEQ), federal (e.g., NOAA), university, and other publicly available sources also facilitate quantitative analysis.

1.4.1. Demographics

The 2020 census reported the population of North Carolina at 10.44 million (U.S. Census Bureau, 2021), up from an estimated 10.15 million in 2016 (U.S. Census Bureau, 2017), and 9.5 million in 2010 (Mackun et al., 2011). It is the 9th most populous state in the U.S. and the 28th largest state in area. North Carolina also provides a setting with rich rural-urban diversity. While North

Carolina's population transitioned to a majority urban state (>50% population in incorporated municipalities) in the 1990s and its urbanization has increased to 57 percent (5.92 million) by 2019, 4.56 million people live in rural areas and the majority of counties (80 of 100) are rural in characteristics, with the majority of the population living in unincorporated areas (Cline, 2020).

1.4.2. Physical geography

North Carolina has some of the most variable topography, physiography, and climate in the eastern United States. Elevation ranges from sea level to 6684 feet (2037 m) at the summit of Mount Mitchell, and is the single largest contributor to the state's temperature variability, which on average differs more than 20°F (~11°C) between the highest mountain elevations and the lower coast in all seasons (State Climate Office of North Carolina, n.d.). North Carolina has three distinct physiographic divisions: the Coastal Plain (Inner Coastal Plain and Outer Coastal Plain, comprising the Tidewater and Outer Banks regions), the Piedmont, and the Blue Ridge Mountains (from east to west) (Raisz, 1940).

Most of the state is classified as a humid subtropical climate (Cfa), with smaller regions in the Appalachian Mountains that have a subtropical highland climate (Cfb) according to the Köppen-Geiger climate classification (Kottek et al., 2006). Average rainfall varies throughout the year in North Carolina, though there are no distinct wet or dry seasons. Summer (July) is normally the wettest season (month) and autumn (November) the driest season (month). Precipitation is heterogeneous across the state with an average annual rainfall of 49.26 in. (range: 36.99, 90.51 in.); the rainiest region in the eastern U.S. is less than 50 miles south of the driest point south of Virginia and east of the Mississippi River (State Climate Office of North Carolina, n.d.).

1.4.3. Water supply, sanitation, and land use exposure routes

Over 40 million U.S. residents (44.5 and 42.5 million in 2010 and 2015 respectively) rely on private groundwater wells (a.k.a. domestic self-supply) as their primary source of drinking water, while 86-87% of the U.S. population is served by public water systems (Dieter et al., 2018; Maupin et al., 2014), which are defined by the US EPA as serving at least 25 persons or having at least 15 service connections (US Environmental Protection Agency (EPA), n.d.). In 2010, North Carolina was second to Pennsylvania (3.35 million; 26%) as having the second highest state self-supplied population (3.30 million; 35%) (Maupin et al., 2014). By 2015, public water supply access and the NC population overall increased leaving 2.4 million (24%) people on private wells. Unlike public water systems, which are federally regulated by the Safe Drinking Water Act and National Primary Drinking Water Regulations (NPDWR) (US Environmental Protection Agency (EPA), 2022), private wells are not under the protection of the Safe Drinking Water Act. Water quality monitoring and treatment is, therefore, optional and the responsibility of the private well owner. Between 2000-2010, for example only 200,000 wells were tested for contaminants in NC, despite the large number of wells (NC DHHS, 2021). Residents reliant on unregulated private wells have increased risk of exposure to microbial and chemical contamination. For example, of the 7.5% (95% CI: 6.6, 7.9%) AGI-related ED visits attributable to microbial contamination of drinking water, DeFelice and colleagues found that 99% (29,200; 95% CI: 26,500, 31,900) were attributable to private wells (DeFelice et al., 2016)).

Areas without public water also typically lack public sewer. Over 50% of the four million occupied homes in North Carolina depend on on-site wastewater and sewage treatment and disposal systems (NC Department of Health and Human Services (NC DHHS), n.d.). Poorly-installed or overloaded septic systems, a shallow or unusually high water table, and flooding can all generate conditions that promote the transport of sewage-derived contaminants into groundwater. Contamination events and waterborne disease incidents have been traced to groundwater wells that were impacted by pathogens from septic drainfields (Gunnarsdottir et al., 2013; Scandura & Sobsey, 1997). Murphy and colleagues (2020) monitored weekly or biweekly samples from 5 private wells in

Pennsylvania and found evidence for human fecal contamination in each well at least once, in addition to significant association between the fecal contamination (total coliforms or human *Bacteroides* HF183) and lagged rainfall. Dye tracer evidence between 3 of the wells and household septic systems in their proximity, which further supports the argument that private well microbial contamination may be attributed to septic systems.

Access to public water supplies is raises environmental justice issues, in addition to health issues, due to racial and socioeconomic disparities in areas that are unserved/underserved by municipal water supplies. In North Carolina, these include peri-urban or donut-hole communities and current or historically Black communities, some of which were denied and have not yet received services since racial segregation was legal (Dewan, 2005; Gilbert, 2013; Johnson et al., 2004; Marsh et al., 2010). DeFelice and colleagues (2016) estimated that if community water services were extended to 10% of the current private well population, approximately 3000 (2,920; 95% CI: 2,650-3,190) annual ED visits could be prevented (Eaves et al., 2022; Flanagan et al., 2016). Reliance on private wells also poses an additional financial burden to test for and treat microbial and/or chemical contamination (Eaves et al., 2022).

Non-point source pollution from agriculture represents another reservoir of fecal organisms that may contaminate water supplies. Potential direct exposure pathways also exist for agricultural workers and other people who come into contact with agricultural waste. After Iowa, North Carolina is the second-largest hog producing state in the country with 9 million hogs (USDA, 2007). The highest density of concentrated animal feeding operations (CAFOs) in the eastern portion of the state. Proximity to industrial swine operations has been associated with respiratory, gastrointestinal, and mucous membrane (nose, throat, eyes) health effects (Wing & Wolf, 2000). Swine/hog production produces large volumes of waste that is stored in lagoons, which may breach or overflow in heavy rains or hurricanes. The manure waste from CAFOs can also be released through sprayfields (Wing et al., 2002). A recent geospatial analysis of stream impairment associated with land use and stormwater management found that swine lagoon density was significantly associated with fecal coliform concentration in N.C. streams (Vitro et al., 2017). Previous studies that examined farming activities, livestock density, and/or manure application as an environmental risk factor for gastrointestinal infections, such as AGI, campylobacteriosis, or giardiasis, have found mixed results depending on the location (Febriani et al., 2010; e.g., Nygård et al., 2004; Odoi, Martin, Michel, Holt, et al., 2004; Odoi, Martin, Michel, Middleton, et al., 2004). However, in a study of AGI ED visits at the ZIP code level in North Carolina (2016-2019), a period that overlapped with Hurricanes Florence (2016) and Matthew (2018), Quist and colleagues (2022) found that AGI rates increased by 11% (95% CI: 1.06, 1.17) in areas with high hog exposure relative to areas without hog exposure, and increased by 21% in rural areas (95% CI: 0.98, 1.43). The association was modified by race/ethnicity (specifically for rural American Indian and Black residents), heavy rain in the prior week, and in areas where swine and poultry CAFOs were co-located. It is clearly plausible that exposure routes through drinking water or environmental contact may be relevant to AGI cases in North Carolina.

This dissertation is the first state-wide study to estimate the relationship between AGI and both temperature and precipitation in North Carolina using high-resolution data, and to model potential effect modification by sociodemographic and environmental factors.

1.5. Specific aims

To improve the understanding of the relationship between weather and diarrheal diseases, this dissertation conducted an original, empirical study of the relationship between acute gastrointestinal illness (AGI) emergency department (ED) visits and short-term changes in weather in North Carolina. It further explored how model specification may affect AGI-weather risk estimates, with particular attention to sensitivity of the effects of rainfall on AGI ED rates to the measurement/definition of "exposure," and to the importance of effect modification by region, water supply type, and industrial hog operations.

This dissertation was structured by the following research objectives.

Aim 1. Review the methodologies used in studies of diarrheal disease and weather.

What methodologies have recent studies adopted to estimate the association between diarrheal diseases and precipitation/weather? What are the strengths and limitations of the approaches they have taken to study design, data sources, outcome, exposure, spatio-temporal scale and resolution, and statistical models and specification? In Chapter 2, we conducted a literature review with explicit consideration of:

- Data sources for outcome and exposure
- Outcome definitions
- Exposure definitions, with an emphasis on rainfall
- Spatio-temporal scale, resolution, and aggregation (unit of analysis), and
- Statistical model type and specification.

Aim 2. Determine the association between precipitation and rates of AGI.

How sensitive is the association between acute gastrointestinal illness (AGI) emergency department (ED) rates and precipitation to the definition or metric used for exposure to precipitation, and, by extension, are differences in measurement/definition likely to contribute to inter-study variability in results? In Chapter 3, we empirically tested various daily precipitation measures from the literature to explore the impact of different meteorological exposure definitions on the apparent relation between AGI and weather. Specifically, we

- used time series analysis (2008-2015) to assess the relationship between precipitation and AGI ED visit rates in NC on the daily, ZIP code level, and

- compared the sensitivity of different daily measures of absolute, extreme, and antecedent precipitation on the association between rainfall and AGI ED visit rates.

Aim 3. Examine the effect measure modification resulting from the variable influence of water supply type and sociodemographics on the association between AGI ED visits and weather.

How do sociodemographic factors and the type of water source modify the relationship between AGI ED rates and weather in North Carolina? In Chapter 4, we expanded on the work of Aim 2 by investigating additional population-level independent variables that may act as effect modifiers to best-fitting model(s). Specifically, we

- examined how the aforementioned relationship between daily AGI ED visits and extreme precipitation in North Carolina by ZIP code (2008-2015) is modified by region, population served by private wells, and industrial hog operations.

2. DIARRHEAL DISEASE AND WEATHER STUDIES: A METHODOLOGICAL LITERATURE REVIEW (AIM 1)

2.1. Introduction

Over the last three decades, researchers have sought to analyze connections between multiple health outcomes and the variables available to represent climate change. Three types of research are needed to meet this challenge (McMichael & Lindgren, 2011): (1) empirical studies on trends in and impacts of climatic variables and on changes in health outcomes that may be climate-sensitive (e.g., risk per °C); (2) statistical studies to estimate the burden of a given health outcome that may be attributed to climate factors; and (3) scenario-based modeling to estimate plausible ranges of future health risks. Many epidemiological studies reviewed here and elsewhere are of the empirical variety; they investigate the effects of climatic factors (e.g., temperature, rainfall, humidity) on general or specific diarrheal diseases, often incorporating statistical techniques (e.g., time series, panel data, spatial models) to estimate the associations between climate variables and health outcomes. However, empirical evidence for many hypothesized connections between meteorological factors and diarrheal diseases is mixed due to the lack of mechanistic understanding between exposure and outcome (Mellor et al., 2016), low availability of data for health outcomes or weather data in many locations for longer time series, challenges in acquiring or limited use of pathogen-specific outcomes (Kraay et al., 2020; K. Levy et al., 2016) or uncertainty in outcome reporting (Lo Iacono et al., 2017), spatiotemporal discrepancies between exposure (e.g., precipitation) measures and diarrheal outcomes can attenuate effect estimates (M. C. Levy et al., 2019), and lack of unified approaches integrate effects at different spatial distributions and time scales (Lo Iacono et al., 2017).

Comparison of results across studies of weather and diarrheal disease or acute gastrointestinal illness (AGI) is complicated by wide variation in metrics of heavy rainfall, outcome definitions, data sources, and analytical methods (Guzman Herrador et al., 2015; K. Levy et al., 2016). Most existing reviews of AGI-climate studies have focused on results, with limited commentary on methods or with limited study coverage (Guzman Herrador et al., 2015; Lo Iacono et al., 2017). Methodological choices in the literature merit explicit examination: they can affect estimates of key results, including regression coefficients and the standard error of the analysis. Methodological differences further contribute to difficulties in understanding whether heterogeneity in results across different studies reflects real differences in climate-health relationships across time or space, or is just a methodological artifact. A notable exception is the methodological systematic review by Lo Iacono and colleagues (2017), which discusses seven key questions regarding: (1) the type and location of water-associated pathogens; (2) methods classification (e.g., descriptive phenomenology, process-based models, empirical statistical analyses); (3) whether the study investigates climate change and/or weather effects; (4) dependence between the method and disease/pathogen; (5) key features of methods; (6) whether and how the results were assessed or method validated; and (7) author-reported methodological limitations; and identifies seven challenges to develop methods to quantifying the effects of weather and climate on water-associated diseases. Our literature review expands on the Lo Iacono *et al.* (2017) by reviewing the modeling approaches and specification decisions used in recent studies investigating the association between weather and AGI to inform the modeling decisions of future research.

We distinguish between two main methodological approaches to studies of climate-health relationships: dynamic mathematical modeling and statistical modeling. Mathematical modeling has primarily been applied to specific diseases like malaria (Beck-Johnson et al., 2013; Pascual, Cazelles, et al., 2008) and cholera (Koelle, 2009; Pascual, Chaves, et al., 2008) and is not the focus of this study. Regression analysis—particularly time series (Bhaskaran et al., 2013; Peng & Dominici, 2008; D. C. Thomas, 2009), case-crossover (Lu & Zeger, 2007; Maclure, 1991; Maclure & Mittleman, 2000), panel, and cohort techniques—is frequently used in environmental epidemiology to study acute health effects or mortality associated with environmental exposures such as temperature (Armstrong, 2006; Gasparrini & Armstrong, 2010) and air pollution (Dominici, 2004; Goldberg et al., 2003; Stieb et al., 2002). Less common are data mining (DM) techniques, which can handle complex relationship and are often used for predictive modeling or forecasting, sometimes with better results than regression analysis (Prasad, Iverson and Liaw, 2006). They include multivariable adaptive regression splines (MARS) (Friedman, 1991), classification and regression tree (CART) or regression tree analysis (RTA) (Breiman et al., 1984; Clark & Pregibon, 1992; Verbyla, 1987), bagged CART (BC) (Hastie et al., 2001, 2009), and random forest (RF) (Breiman, 2001). With few notable exceptions, the majority of statistical studies of climate-health relationships report findings from one or perhaps two types of models without comparing impacts of model selection on study conclusions.

To facilitate the interpretation and use of existing data related to diarrheal disease and climate, we undertook an extensive methodological review of the literature with explicit focus on data sources for outcome and exposure; outcome definitions; exposure definitions, with an emphasis on rainfall; spatio-temporal scale, resolution, and aggregation (unit of analysis); and statistical model type and specification, focusing on time series and case-crossover methods with passing attention to data mining techniques, which were not well-represented in the reviewed studies. We sought to avoid the lack of overlap exemplified by two earlier reviews of research modeling associations between diarrheal diseases and precipitation (Guzman Herrador et al., 2015; K. Levy et al., 2016): these reviews examined 37 case-based (i.e., non-outbreak) studies between them, yet only two of the studies appeared in more than one of the review papers. To this end, we compiled results from four more recent systematic reviews (published between 2020-2022) and sought out a range of up-to-date modeling methods.

Our review focused on a comparison of methodological practice and on how model specifications affect results. We examined sources of heterogeneity across diarrheal disease and rainfall studies, as well as the study characteristics and methods employed in historical empirical research, in the hopes of informing study and statistical practices to increase comparability across future studies. We begin with an overview of the different statistical approaches and health outcomes that have been used to model historic AGI-weather relationships. Then we review how studies that model the relationships between climatic factors and diarrheal diseases have handled a range of modeling challenges and discuss the strengths and limitations of these methods.

2.2. Methods

We used studies included in the meta-analyses of four recent (2020-2022) systematic reviews of statistical studies on the relationship between diarrheal diseases and weather exposures (excluding seasonality, floods, droughts, and storms) with either temperature and/or precipitation as one of the exposures (Chua et al., 2021, 2022; Kraay et al., 2020; M. Liang et al., 2021; P. Wang et al., 2021). We selected these recent systematic reviews for because they represented some of the most recent literature on the relationship between diarrheal diseases and weather and because of a recently expanded systematic search (Kraay et al., 2020) on precipitation and diarrhea studies since Levy *et al.* (2016).

The process of study identification, screening, and inclusion is diagrammed in FIGURE 1. These four studies focused on diarrheal diseases and temperature (Chua et al., 2021, 2022; M. Liang et al., 2021), diarrheal diseases and rainfall (Kraay et al., 2020), or lagged associations in climate-health studies, yielding a total of 234 studies. The studies in Wang *et al.* (2021) included multiple health outcomes and, therefore, were subsetted to include only the 39 diarrhea, salmonella, and cholera studies. We retained the 16 all-cause diarrheal studies that were dropped between the pre-print version (Chua et al., 2021) and the final publication of the systematic review and meta-analysis by Chua and colleagues (2022), which included only pathogen-specific studies. Between the two publications (Chua et al., 2021, 2022), 80 of the studies were included in the narrative review, 56 in the pre-print meta-analysis (2021), and 40 in the final peer-reviewed publication (2022). We identified an additional 88 studies from Kraay *et al.* (2020) and 27 studies from Liang *et al.* (2021). We extracted the studies and relevant tables from each of the systematic reviews, recorded the source(s) of the articles, and deduplicated the studies, resulting in the removal of 52 studies, for a total of 182 records. These were screened to include only studies that had been included in any of the meta-analyses, excluding studies that were only included in narrative reviews (n=131).

Eligibility criteria for inclusion in the methodological review included: journal articles in the English language only; full texts available; use of any time series statistical model or case-crossover methods to estimate the associations between climatic factors and all-cause or pathogen-specific diarrheal diseases or acute gastrointestinal illness (AGI), excluding studies that only included mathematical models and correlation analyses without regression models; and cases only, excluding outbreaks. We included studies in which one of the main exposure variables was temperature or precipitation, and excluded studies that primarily models the association between flooding, storms/hurricanes/extreme events, seasonality, and ENSO only. Compared to many of the other pathogens with diarrheal symptoms, cholera is a special case in which there are more studies that include more mathematical modeling methods (e.g., Bouma & Pascual, 2001; Koelle, 2009; Koelle et al., 2005; Koelle & Pascual, 2004; Pascual et al., 2000, 2011; Rodó et al., 2002), which are outside the scope of this study; cholera was included if the study used a time series or case-crossover model. All studies included the source systematic reviews were already limited to epidemiological studies of human disease and did not include studies of animal infections or microbial water quality studies.

We treated the following terms for diarrheal or enteric illnesses used by different studies and reviews synonymously—(infectious) diarrheal disease, waterborne diseases, enteric illnesses, (acute) gastrointestinal illness (GI or AGI), gastroenteritis, etc.—and distinguished between grouped all-cause and pathogen-specific illnesses.

As the focus of this review is methodological and the source systematic reviews have reported on their results, we extracted information most pertinent to methods and modeling. The following 13 categories of data were ultimately extracted from the studies: outcome etiology; study setting (country, sub-national location); outcome measure; study population; outcome data source; spatial resolution for weather exposure; model spatial unit of analysis; study period; temporal study design; model time step; statistical methods; weather exposure; covariates. One reviewer extracted relevant data columns and tables from the published systematic reviews, matched them with the aforementioned categories of information (as applicable) and merged the tables together. After merging the data tables from each of the studies, we edited the information to match the data to be extracted and identified and filled any gaps from the studies to develop a large, comprehensive data table of the studies. The analysis and results of this review, including model references and examples, are summarized in TABLE 4 and TABLE 5.

2.3. Model specification elements

Important study characteristics and elements of model specification include the following: a) etiology (all-cause or specific disease) and outcome measure; b) age of the population of interest; c) outcome data source; model unit of analysis; d) time period of the analysis; e) country and f) location of the study setting; g) spatial coverage; h) model unit-of-analysis (UOA); i) model time step and study type with respect to time; j) type(s) of statistical model(s); k) sources of weather data; l) meteorological variables modeled; m) covariates reflecting the level of control for other confounding variables or effect modifiers. Any and all of these factors may vary from one study to the next, making it difficult distinguish whether observed differences of outcomes reflect real differences or result from study design differences (Butler and Hall, 2009). Below, we comment in more detail on patterns across studies that are presented in TABLE 4 (A-H) and TABLE 5 (A, I-M). Counts (n) are presented out of the 98 studies for sections A-M, unless otherwise noted.

2.4. Statistical model types

2.4.1. Statistical model types (TABLE 5 Column J)

The vast majority of empirical studies on climate and health in the environmental epidemiological literature apply regression models to time series variation to assess the historical impacts of the seasonality, variability, or extremes of climatic factors on disease or seasonality of disease. Researchers have used a number of regression models to investigate the influence of weather on acute gastrointestinal illness (AGI) and diarrheal diseases. Time series generalized linear models (GLM), generalized additive models (GAM), and generalized estimating equations (GEE), in addition to case-crossover models, are the most common. Some studies use count or rate models—with Poisson, quasi-Poisson, negative binomial, or log-linear transformations—for outcomes like hospital admissions, cases, incidence, or deaths at a daily, weekly, or monthly time-scale. To better account for seasonal disease dynamics, researchers may control for seasonality using time-stratified variables (e.g., month), Fourier terms, or splines, or employ autoregressive moving average (ARMA) (Drayna et al., 2010) or seasonal autoregressive integrated moving average (SARIMA) models (Y. Zhang et al., 2007, 2008a, 2010).

As presented in TABLE 5 Column J and summarized in SI TABLE 11, the most common statistical models in the review studies were time series generalized linear models (GLM): Poisson GLMs (n = 47),[6, 8, 10, 11, 14, 16-19, 26-34, 37, 38, 41-44, 46-50, 53, 57, 59, 60, 62, 65, 71, 74-77, 81-83, 85, 89, 90, 92, 98] some of which accounted for over-dispersion using quasi-Poisson transformations [e.g., 10, 14, 16, 17, 18, 26, 34, 38, 46, 77, 83, 85, 92] or were multilevel models[e.g., 82]; Poisson, logistic, or log-binomial generalized linear mixed models

(GLMM) (n=7), [5, 8, 25, 27, 32, 67, 82] which include random effects terms; and negative binomial regression models (NBRM) [1, 2, 4, 24, 40, 55, 61, 63, 64, 69, 70, 72, 73, 79, 83, 86, 87, 88] and zero-inflated NBRMs[58] (n=19). Generalized additive models (GAM) were the second most common: Poisson or quasi-Poisson GAMs (n=9),[7, 9, 13, 22, 39, 42, 87, 91, 93] negative binomial (NB) GAMs (n=2),[61, 78] and log-binomial generalized additive mixed models (GAMM) (n=1).[25] Less frequent models included ordinary least squares (OLS) (n=2),[20, 63, 20] multiple linear regression (n=3),[21, 31, 52, 72] logistic regression (n=8),[3, 5, 7, 12, 15, 36, 47, 56] log-linear models (n=3),[23, 64, 66] generalized least squares (n=1),[68] and Poisson generalized estimating equations (GEE) (n=2) (proposed by K.-Y. Liang & Zeger, 1986; Zeger & Liang, 1986).[59, 68] Ten studies included random effects predictors in addition to fixed effects of various types (e.g., GLMM, GAMM, NBRM, GLS, Bayesian Poisson model) (n=10).[5, 8, 25, 27, 32, 35, 67, 68, 82, 86] Distributed lag terms for meteorological exposure variables were incorporated into GLM(M)s and GAM(M)s with (linear) distributed lag models (DLM) (n=13)[7, 16, 17, 18, 28, 29, 46, 48, 49, 62, 69, 70, 74] (see Schwartz, 2000b) or distributed lag non-linear models (DLNM) (n=18)[10, 14, 26, 33, 34, 37, 38, 47, 57, 61, 75, 77, 78, 83, 85, 89, 91, 92] (see Gasparrini et al., 2010). Three studies employed Bayesian statistics (see review by van de Schoot et al., 2021) in Bayesian Poisson models (n=1)[35] or Bayesian Space-time Hierarchical Models (BSTHM) (Wikle et al., 1998) (n=2).[94, 97] A minority (n=7) of time series studies used auto-regressive (AR)[51] or (Seasonal) Auto-Regressive Integrated Moving Average (with eXogenous factors), or (S)ARIMA(X),[45, 54, 80, 84, 95, 96] models. Case-crossover models (Maclure, 1991; Maclure & Mittleman, 2000) were used in four studies.[15, 44, 47, 56] Eight studies used meta-analysis or meta-regression methods (n=8).[23, 34, 59, 60, 62, 67, 68, 85] Common types of statistical model with references and R packages are presented in TABLE 6.

Most studies reported the findings from a single type of model, rather than comparing the impacts of different of regression models on study conclusions, except 18 studies that used at least two types of models (n=18),[7, 15, 25, 42, 44, 47, 59, 61-64, 68, 72, 73, 83, 87, 91, 97] most of which compared two models from amongst (quasi-)Poisson GLMs, NBRMs, and (quasi-)Poisson GAMs with or without distributed (non-linear) lags. White and colleagues (2009)[44] compared Poisson GLM and case-crossover models, while and Eisenberg and colleagues (2013)[47] compared two statistical models—Poisson GLM with DLNM and case-crossover—to a dynamic susceptible-infectious-water-recovered (SIWR) model (Tien & Earn, 2010). Eisenberg and colleagues (2013)[47] found agreement between the three models with a strong relationship between rainfall and cholera at the regional and country scales. Zhang and colleagues (2008a) and Shortridge and Guikema (2014) are amongst the few other studies have explicitly compared model types for diarrheal diseases and environmental exposures. Zhang and colleagues (2008a) compared four regression models—standard Poisson regression, autoregressive adjusted Poisson regression, multiple linear regression, and a SARIMA—in a study of the association between climate variation and salmonellosis in Australia. They found that the SARIMA model performed best based on goodness-of-fit and forecasting ability. Shortridge and Guikema (2014) compared the performance of two regression models (Poisson GLM and Poisson GAM model) with four data mining models—multivariate adaptive regression splines (MARS), classification and regression tree (CART), bagged CART, and random forest (RF)—to investigate the relationship between gastrointestinal illness and pipe breaks in two American cities. The authors found that the RF and bagged CART performed the best according to their criteria of in-sample and out-of-sample accuracy. Multi-model approaches have been recommended as they decrease the likelihood of misdrawing conclusions due to the artifacts of model-specific assumptions and examine how complementary information may be learned when comparing different tools (M. C. Eisenberg et al., 2013).

The following subsections examine the most important statistical model types in more detail: Poisson-transformed GLM, Poisson-transformed GAM, distributed lag (non-linear) models, (seasonal) autoregressive integrated moving average (with exogenous variables), and case-crossover analyses.

Time series (TS) models

In the last decade or so, several critical reviews have been published about time series models in environmental health. Grasso and colleagues reviewed time-series models as well as other quantitative methods—including panel and spatial models, and non-statistical approaches such as the integrated assessment models (IAMs)—used in studies on the health effects of climate change (Grasso et al., 2012). Other publications (Bhaskaran et al., 2013; Gasparrini & Armstrong, 2010; Imai et al., 2015) and systematic reviews (Imai & Hashizume, 2015) have discussed and provide useful advice on potential approaches to methodological issues with time series regression models applied to environmental epidemiology for infectious disease and environmental and climatic factors. Here, we summarize four major approaches to time-series modeling—Poisson GLM, Poisson GAM, DLM/DLNM, (S)ARIMA(X)--with the exception of generalized estimating equations (GEE) for clustered data (refer to Agresti, 2002, secs. 11.3-11.4; Nitta et al., 2010).

TS Model 1: Poisson-transformed generalized linear model

Used in 48% of the reviewed studies, the Poisson-transformed generalized linear model (GLM) (a.k.a. Poisson GLM or GLM with Poisson error structure) is a maximum likelihood estimation (MLE) method that models count data by using a random component with a Poisson distribution and a log link function (Cameron & Trivedi, 1998, 2013; McCullagh & Nelder, 1989). With the addition of a population offset, it can be used to model rates. This method can use a mix of continuous and categorical explanatory variables as system components. When all explanatory variables are discrete, the log-linear (GLM) model is equivalent to the Poisson regression model and the Poisson regression model for counts is sometimes referred to as the "Poisson log-linear model" (Armstrong, 2006; Bhaskaran et al., 2013).

The Poisson regression model assumes the variance function ($var(Y_i)$) is proportional to the mean (μ_i): $var(Y_i) = \tau\mu_i$, where τ, the dispersion parameter, is equal to one. When there is evidence of overdispersion (i.e., when variance of the dependent variable is larger than the mean), several options exist. The model can be adjusted for by using a quasi-Poisson (a.k.a. Poisson with overdispersion) model, which is a quasi-likelihood model that fits the dispersion parameter to the data to account for the extra variance. Alternately, a negative binomial model (NBRM) may be used to handle overdispersion more formally (Butler and Hall, 2009; Guzman Herrador et al., 2015; Carlton et al., 2016). Hurdle and zero-inflation Poisson (ZIP) models may be used to model excess zeros. The standard Poisson model assumes that outcome variables are independent given covariates and does not control for autocorrelation. If assumptions for serial independence are violated—as in the case of high person-to-person transmission from infectious to susceptible individuals—it could underestimate the standard errors of the estimates. Therefore, autoregressive-adjusted Poisson regression may be used to account for serial autocorrelation in the outcome by including lagged outcome terms as predictors (e.g., Y_{t-1}). GLMs assume a linear relationship between the outcome and exposure, which could be violated if there are non-linearities in the relationship between the diarrheal disease outcome and meteorological exposure variable. Non-linearities may be explored further by relaxing linearity assumptions with generalized additive models (GAM) or with distributed lag non-linear models (DLNM). Additional extensions included mixed effects models (GLMM), multilevel models to allow for nested spatial data (Finch et al., 2019; Gelman & Hill, 2008), and two-stage models that incorporate observation-level random effects.[e.g., 6, 7, 34, 85]

TS Model 2: Poisson-transformed generalized additive model

Generalized additive models (GAMs) are used to model non-linear relationships between outcome variable and covariates through the use of smoothing functions, originally developed by Hastie and Tibshirani (1986) (refer to Hastie & Tibshirani, 1986, 1990a; Wood, 2004, 2006, 2017).

GAMs use three types of smoothers: regression splines (B-spline, P-spline), local regression (loess), or smoothing splines that minimize the penalized sum of squares (Ravindra et al., 2019), and are robust to the selection of the smoothing spline (Peng, Dominici and Louis, 2006). Penalized likelihood maximization is used to fit smoothing functions, and penalties can be set as low as zero for covariates that do not improve the model (Hastie & Tibshirani, 1990a). Siilarly to GLMs, GAMs can be (quasi-)Poisson or negative binomial-transformed or be used in mixed effects (GAMM) and multilevel models. In AGI-weather studies, GAM terms may be applied to exposure variable(s) to represent non-linear relationships that have been observed between AGI and climate factors, such as precipitation and temperature.[7, 9, 13, 22, 25, 39, 42, 61, 78, 87, 91, 93] GAMs are further discussed in the review article by Ravindra and colleagues (2019) in the context of air pollution, climatic variability, and health outcomes, but many of their methodological insights are applicable to AGI-weather research.

TS Model 3: Distributed Lag Models (DLM) and Distributed Lag Non-linear Models (DLNM)

Distributed lag models (DLMs) and distributed lag non-linear models (DLNMs) (Gasparrini, 2011, 2014; Gasparrini et al., 2010, 2017) address the challenge of selecting and modeling lagged independent or exposure variables in time series analysis and can be combined with various models (e.g., GLM, GAM). Rather than using a single lag or aggregating lags over a set amount of time, DLMs allocate the delayed effect of a single exposure event over a period of time in an attempt to represent the time course of the exposure-response relationship (Gasparrini, 2011). They can describe the exposure-response relationship in time either forwards (fixed exposure to future outcomes) or backwards (fixed outcome to past exposures). DLMs were originally developed by Almon (1965) for the field of econometrics. Over the last two decades, they have been adapted to studies on the effects of environmental factors on health, starting with air pollution (Muggeo & Hajat, 2009a; Roberts & Martin, 2007; Schwartz, 2000a; Zanobetti et al., 2000) and temperature (Ferreira Braga, Zanobetti and Schwartz, 2001) and have been incorporated in infectious disease and weather studies, including distributed lag terms for temperature and precipitation AGI-weather studies (DLMs[7, 16, 17, 18, 28, 29, 46, 48, 49, 62, 69, 70, 74] and DLNMs[10, 14, 26, 33, 34, 37, 38, 47, 57, 61, 75, 77, 78, 83, 85, 89, 91, 92]).

Distributed lag non-linear models (DLNM) relax the assumptions of the DLM regarding the shape of the exposure-response relationship to account for non-linearities (Gasparrini, Armstrong and Kenward, 2010). They flexibly describe "effects that vary simultaneously along the space of the predictor and in the lag dimension of its occurrence" (Gasparrini, Armstrong and Kenward, 2010). The most common DLNM methods currently in use were originally proposed by Armstrong (2006c), further developed and applied in R using the *dlnm* package by Gasparrini, Armstrong, and colleagues (Gasparrini, 2011; Gasparrini et al., 2010, 2017; Gasparrini & Leone, 2014), and were then extended to exposure-lag-response associations (Gasparrini, 2014).

The development of a DLNM involves three main steps: (1) define the exposure-response relationship in the space of the predictor using basis functions (e.g., natural cubic or B-splines, dummy variables, polynomials, or threshold-type) to transform the original variable into a new set of basis variables; (2) specify the function to model the additional lag dimension; and (3) simultaneously define the exposure-response relationship by selecting a bi-dimensional cross-basis for predictor and lags (Gasparrini, 2011). Gasparrini (2014) has expanded the aforementioned three main steps into nine practical DLNM analysis steps to follow when using the *dlnm* R package.

TS Model 4: (S)ARIMA(X) – (Seasonal) Autoregressive Integrated Moving Average models (with eXogenous variables)

Time series often contain trends and seasonal patterns that are non-stationary in nature (Adhikari & Agrawal, 2013). For a stochastic process to be stationary, its statistical properties—mean, variance, and autocorrelation structure—do not change over time. However, a data series may exhibit non-stationarity in the form of a time-dependent mean or variance, or a periodic or seasonal component (Gardner, 1985, chap. 12; Gardner, Napolitano and Paura, 2006; Napolitano, 2016) (Chatfield, 2001, chap. 3; Cryer and Chan, 2008, chap. 2). A time series can be made stationary through differencing and power transformations to remove the trend and/or seasonal components. Due to the seasonality of AGI, models that decompose the outcome signals into seasonal, trend, and residual components through stochastic process modeling are likely to be useful.

Stochastic process models include and build from autoregressive, AR(p), and moving average, MA(q), models, which can be combined into other models (Chatfield, 2001; Shumway and Stoffer, 2006, 2017; Cryer and Chan, 2008):

- ARMA(p,q) – autoregressive moving average
- ARIMA(p,d,q) – autoregressive integrated moving average
- SARIMA(p,d,q)$(P,D,Q)_s$ – seasonal autoregressive integrated moving average
- ARIMAX(p,d,q) – autoregressive integrated moving average with exogenous variables
- SARIMAX(p,d,q)$(P,D,Q)_s$ – seasonal autoregressive integrated moving average with exogenous variables

where the variables *p*, *d*, *q*, and *s* are integers greater or equal to zero, defined as below (Adhikari and Agrawal, 2013):

p	non-seasonal autoregressive (AR) order
d	non-seasonal differencing order (I) (integrated process of order d (I(d))
q	non-seasonal moving average (MA) order
P	seasonal autoregressive (AR) order
D	seasonal differencing order (integrated process of order D (I(D))
Q	seasonal moving average (MA) order
s	periodic term (length of seasonal period)

ARMA models are only suitable for stationary processes, while ARIMA and SARIMA models can account for the non-stationary processes by incorporating differencing to remove the trend (d) and/or seasonal (D) components (Chatfield, 2001; Shumway and Stoffer, 2006, 2017; Cryer and Chan, 2008).

ARIMA models, the most general, have three main advantages: (1) the relationship between the current state as a function of both endogenous variables and exogenous variables is easily interpretable for retrospective studies; (2) model selection over time series can be automated to maximize prediction accuracy, and (3) the ARIMA model handle dynamic relationships over time by updating the model based on recent events for future predictions (Kane et al., 2014). The disadvantages of the ARIMA model stem from two of its assumptions: (1) relationships between independent and dependent variables are assumed linear; and (2) standard deviation in errors in the model over time are assumed constant (Kane et al., 2014). Seasonal ARIMA, or SARIMA models, have been used more often than ARIMA models in AGI-weather studies and may be more accurate than ARIMA for AGI time series because they account for seasonality, but have been shown to be sensitive to the periodic term *s* (Valipour, 2015). ARIMAX and SARIMAX models are modifications of (S)ARIMA models that include exogenous covariates (X), such as meteorological factors, depending on the cross-relations between the exogenous and response variables (H. S. Lee et al., 2013). (S)ARIMA(X) models were used in six of the reviewed AGI-weather studies,[45, 54, 80, 84, 95, 96] but are not applicable to multiple time series (i.e., for time series at different spatial locations). In the case of multiple time series, vector autoregression (VAR) (Shumway & Stoffer, 2017) or Bayesian VAR may be considered as alternatives (Karlsson, 2013), though they are more commonly used for forecasting.

Case-crossover analyses

Case-crossover analysis compares the same person—the "unit of observation"—at different periods of time. Originally developed to study the health effects of air pollution (see Carracedo-Martínez et al., 2010), case-crossover analysis is best applied to intermittent exposures with transient acute effects and short induction times (Mittleman and Mostofsky, 2014). It controls for time-invariant individual, seasonal, and geographic differences by matching the exposure of an individual during a case-defining event to a control period for the same individual (Maclure, 1991; Mittleman, Maclure and Robins, 1995; Maclure and Mittleman, 2000). The key assumption is that neither exposure nor confounding variables change systematically during the study period. Case-crossover design with unidirectional sampling was first proposed by Maclure (1991) and has since been modified to decrease confounding bias from time trends in the exposure (Carracedo-Martínez et al., 2010; Perrakis et al., 2014) using full-stratum bidirectional (Navidi, 1998), symmetric bidirectional (SBI) (Bateson and Schwartz, 1999), semisymmetric bidirectional (Navidi and Weinhandl, 2002), and time-stratified (TS) models (Lumley and Levy, 2000).

More recently, case-crossover models have been applied to studies of AGI outbreaks and rainfall (K. M. Thomas et al., 2006; Nichols et al., 2009); AGI cases and flooding (Ding et al., 2013; Lin, Wade and Hilborn, 2015), rainfall (Eisenberg et al., 2013; Gleason & Fagliano, 2017),[15, 47] and sanitary sewer overflows (Jagai et al., 2017); campylobacteriosis and temperature and relative humidity (White et al., 2009);[44] and cholera and heatwaves with effect modification by rainfall and tree cover (J. Wu et al., 2018).[56] Most of these studies have defined the case by a distinct outcome event (e.g., outbreak date or healthcare admission date) or exposure event (e.g., flood, sanitary sewer overflow). For example, Gleason and Fagliano (2017), studied the association between in-patient hospitalizations for gastrointestinal illness and heavy rainfall in New Jersey (2009-2013) using a time-stratified bidirectional case-crossover design and conditional logistic regression with stratification by season (cold/warm) and drinking water source (groundwater/surface water/other). The authors defined the case day as the date of hospital admission and selected two control days from days that (a) occurred on the same day of the week as the case day, and (b) shared the same, fixed 21-day stratum as the case in weeks before or after the case day. Relative odds of exposure to heavy rainfall ($\geq 90^{th}$ percentile for same day or 3-day average rainfall) were estimated between case and control days, and the models were controlled for same day temperature and relative humidity.

2.4.2. Outcome data sources

Sources of outcome data (TABLE 4 Column C): [update values, incorporate, revise]

A number of approaches exist for collecting outcome data on diseases (TABLE 4 Column C). Most studies obtained data from health system, hospital, or other passive surveillance systems (n=59)[1, 2, 4, 10, 11, 13-17, 20, 24, 26, 27, 29, 31, 33-35, 40-44, 48, 54, 57, 60, 62, 65, 66, 69-76, 80-87, 88-95, 97] or from health facilities (e.g., hospitals, emergency departments, and/or outpatient clinics) (n=21),[6, 7, 9, 18, 21, 22, 28, 32, 37-39, 47; 51, 58, 61, 63, 64, 77, 78, 96, 98] which may be used in ecological or case-control studies. After data obtained from health systems, community-based outcome data (n=8)[3, 5, 8, 12, 19, 25, 30, 36] was the next most common; this data is often collected at the household-level via active community- or population-based surveillance (e.g., regular surveys, such as every two weeks); cross-sectional (e.g., Demographic and Health Surveys, DHS), panel, or cohort surveys. Four studies obtained hospital and/or community-based data from published studies to use in meta-analyses or meta-regressions (n=4).[23, 59, 67, 68] Other data sources included national health datasets (n=3),[47, 55, 86] government reports (n=1),[79] epidemiological case reports on cholera[52] or global health records[53] on cholera from the World Health Organization (WHO) (n = 2), outbreak surveillance from NGO medical registries (n=1),[50] and internally displaced persons (IDP) camps (n=1).[47] While most of the studies reviewed used ecological study designs with time series (n=88)[1, 2, 4, 6-11, 13, 14, 16-18, 20-24, 26-29, 31-35, 37-55, 57-98] or case-crossover models (n=4),[15, 44, 47, 56] they also included cross-sectional (n=3),[5, 12, 31] case-control (n=1),[1,3] and cohort study designs (n=5).[3, 19, 25, 30, 36] In their systematic review of the methods used by gastrointestinal illness and drinking water studies, Bylund and colleagues (2017) describe and discuss the advantages and limitations of different data sources for gastrointestinal illnesses—health care (H), patient registers (R), telephone triage (T), pharmacies (P), questionnaires (Q), interviews (I), and health diaries (D—that are frequently used in similar ecological (H, P, T), case-control (H), cross-sectional (Q, I), cohort (Q, I, D), and household intervention or experimental (R, Q, I, D) study designs.

The source and type of outcome data can affect data quality, collection frequency, and severity of observed cases. Outcomes in different data sets may be defined and measured differently. For example, laboratory diagnoses are accurate but expensive, while 1- to 2-week recall of diarrhea is less precise. Billing diagnosis codes from the 9th or 10th Revision of the International Classification of Diseases–Clinical Modification (ICD-9-CM or ICD-10-CM) from hospitals or emergency departments (ED) may be less expensive and easier to obtain from public datasets, but also present challenges; they may not be based on laboratory testing and may underreport the pathogen-specific diagnoses of specific enteric due to lack of or incorrect reported even when an illness is culture-confirmed (Scallan et al., 2018), acute gastrointestinal illness ED visits likely underestimate total AGI incidence (Mead et al., 1999), and challenges mapping between ICD-9-CM and ICD-10-CM codes due to the switch in October 2015 (Krive et al., 2015). Furthermore, health facility outcome data is likely to have more frequent time series information and may be more publicly accessible, but depending on the dataset and on privacy limitations, other desirable information for environmental health studies often goes uncollected or is unavailable to researches. These desirable data include detailed location or spatial information; demographic and socioeconomic details; and information about water supply and sanitation details that could be used as control variables or effect modifiers in models and are likely to be relevant in understanding the interactions between risk factors of diarrheal diseases. Surveys vary in the frequency of data collection and their spatial representation, are expensive to conduct well, and, with some exceptions, are often not publicly available.

Survey data with health outcomes are advantageous because they are more likely to have information relevant to risk factors (i.e., demographic, spatial, environmental information). However, surveys are expensive to conduct well and the results often are not publicly available. The frequency and spatial representation of data collection varies across surveys, and temporal signals that can be gained from frequent case or incidence measurements may be lost if surveys are cross-sectional or if there are long periods of time between panel data collection. Unfortunately, such temporal signals can be important in environmental health research due to the seasonality of diarrheal illnesses. Some studies attempt to compensate for a lack of time series by reconstructing antecedent weather based on the date of survey (Bandyopadhyay et al., 2012).

2.4.3. Outcome definitions and etiologies

Etiologies of diarrheal outcomes (TABLE 4 Column A): [update, incorporate, revise]

There is substantial variation in the degree of disease specificity and the etiologies represented in the outcomes used across studies in this literature (TABLE 4 Column A and summarized in SI TABLE 8). The largest number of studies consider aggregated outcomes of *all-cause* gastroenteritis that cluster together multiple underlying pathogens, including all-cause diarrhea (e.g., unspecified, infectious, or non-cholera), all-cause infectious gastroenteritis (GI) or acute gastrointestinal illness (AGI), or food- and/or waterborne diseases (n=39).[1-39] Many studies choose to focus on disease caused by specific bacterial, protozoan, or viral pathogens if pathogen-specific or laboratory confirmed data is available. The latter approach can be justified by the fact that all pathogens do not respond to environmental conditions (of which weather and climate factors are only one aspect) in the same way. Of the 98 reviewed studies, the most studied pathogens in the literature have been shigellosis (a.k.a. bacillary dysentery) (n=16),[82-97] salmonellosis (n=15),[20, 42, 69-81] cholera (n=12),[45-56] rotavirus (n=8),[61- 68] and campylobacteriosis (n=6).[45-56] Other studies have considered enteritis attributed to *Escherichia coli (E. coli)* (n=3),[42, 58, 59] norovirus (n=2),[60, 61] typhoid fever (n=2),[77, 98] and an aggregated outcome for the protozoa cryptosporidiosis and giardiasis (n=1).[57] Four of the studies reported multiple outcomes (2-3 pathogen-specific and/or all-cause outcomes), which are listed in separate rows by etiology.[20, 42, 61, 77] The outcome measures that were analyzed varied between cases (including health facility visits or admissions), incidence and prevalence, though all studies were focused on diarrheal morbidity rather than mortality (TABLE 4

Column A). One meta-analysis used normalized outcome measures (z-scores) to explain deviations from the norm within the various study sites.[67] More details on specific food- and/or waterborne pathogens and organisms and their climate sensitivities may be found in prior literature reviews (Lo Iacono et al., 2017; Semenza, Herbst, et al., 2012; Sterk et al., 2013). In their systematic review, Lo Iacono and colleagues detailed the transmission routes, organism types (e.g., bacteria, cyanobacteria, virus, protozoan, flatworm, roundworm, fungus, dinoflagellate, diatom), relationship to water, and neglected tropical disease classification of 120 non-vector-borne organisms and the proportion of reviewed papers by pathogen (Lo Iacono et al., 2017, pp. 5–11), and discuss data and methodological challenges of quantifying weather and climate effects on water-associated diseases. Semenza and colleagues reviewed the epidemiology, seasonality, and water, food, temperature, climatic, and environmental determinants of *Campylobacter*, *Cryptosporidium*, *Listeria*, norovirus, *Salmonella*, and *Vibrio* (Semenza, Herbst, et al., 2012). Sterk and colleagues systematically reviewed studies of the pathogens *Campylobacter*, *Cryptosporidium*, norovirus, and *Vibrio* in relation the climate change effects of risk of infection using a conceptual model based on quantitative microbial risk assessment (QMRA) steps and discussed hazard identification, input sources (e.g., human, animal), and compartments (land surface, surface waters, sediments, and aquifer) by pathogen (Sterk et al., 2013).

Age groups of study population (TABLE 4 Column B):

Many studies in the epidemiological literature on diarrheal disease focus specifically on children (especially under 5), due to the higher prevalence of diarrheal disease in this population and concomitant improvement in statistical power. However, climate and diarrhea studies, perhaps understanding the importance of having population-based estimates, have more often considered single (n=70) or multiple age groups (n=18)[11, 13, 15, 16, 18, 23, 26, 31, 33, 35, 38, 56, 68, 83, 85, 89, 92, 93] composed of all ages (n=75)[3, 4, 7, 8, 10, 11, 14-17, 19, 20, 22-24, 26, 28-31, 33-35, 39-50, 52-60, 65, 67-84, 86-98] compared to using disaggregated age groups with various age classifications of: infants (e.g., ≤2 years), children (e.g., <5, <6, <10), and/or youth (e.g., <15, <16) (n=32),[1, 2, 9, 11-13, 15, 16, 18, 23, 25-27, 32, 33, 35-38, 51, 56, 61-64, 66, 68, 83, 85, 89, 92, 93] adults (n=11),[5, 11, 15, 16, 18, 21, 26, 33, 56, 83, 93] and/or the elderly (≥65) (n=7)[5, 11, 15, 16, 18, 21, 26, 33, 56, 83, 93] (TABLE 4 Column B and summarized in SI TABLE 10). If age-related data were available, a number of studies considered multiple age groups (n=18),[11, 13, 15, 16, 18, 23, 26, 31, 33, 35, 38, 56, 68, 83, 85, 89, 92, 93] comparing either between age-disaggregated models only or pooled all-age models to age-disaggregated models. Few studies disaggregated results by gender, including male and female populations in addition to aggregate estimates [e.g., 20, 55, 69, 93] (results not reported). Age-disaggregated models provide additional insights to at-risk populations, but using population-based results in climate change projections poses certain challenges, as it requires an assumption that the age distribution in different regions is stable over time or else requires disaggregation of disease outcomes across different age groups. But only a minority of studies report such disaggregated results without reporting results for all ages.

2.4.4. Meteorological data sources

Sources of weather data (TABLE 5 Column K):

There are three main sources of data on historical weather: weather stations, gridded climate or weather data (GCD or GWD) products, and climate data assimilation (DA) or 'reanalysis' products. Each source of weather data has its own strengths and weaknesses, which are summarized from Auffhammer and colleagues (2013) in TABLE 7, supplemented with well-known datasets. GCD and climate reanalysis products were been developed in part to fill some of the "holes" in space and time where station-level weather observations are sparse. They are presented as gridded data and may use different methods, such as statistical interpolation or modeling techniques, to integrate data from ground- and satellite-based observations. Given the spatial variability in diarrheal disease, epidemiologists generally prefer to use as fine a resolution as possible for weather data (Kolstad & Johansson, 2011), which necessitates using data sources depending on local, regional, or global data availability. Consistently, the relevant weather variables for site-specific studies were most often measured at the local level using data from single or multiple (aggregated) weather stations (n=82)[1, 4, 6, 7, 9-11, 13-16, 18-22, 24-30, 32-37, 39- 42, 43-45, 47-51, 54-56, 58-69, 71- 85, 87- 92, 94-98] and rain gauges (n=5)[3, 4, 8, 25, 47] or temperature probes n=1)[4] installed by researchers (TABLE 5 Column K). Historic weather station data was derived from local, national (e.g., from the China Meteorological Administration (CMA), including the China National Meteorological Information Center or China Meteorological Data Sharing Service System, CMDSS, (http://data.cma.cn/)[13, 33, 34, 84, 85, 90, 91, 92, 94, 97]), and global (e.g., NOAA National Climatic Data Center, NCDC[10, 18, 46, 56]; Global Historic Climatology Network, GHCN[59]) weather station datasets or meteorological organizations. Fifteen studies used weather data products that incorporated satellite observations or other data: GCD products (n=13),[2, 5, 12, 17, 23, 46, 47, 52, 53, 57, 59, 70, 93] climate reanalysis products (n=6),[2, 12, 17, 31, 38, 53] and land data assimilation systems (LDAS) (n=2),[2, 38] which are a type of reanalysis product that uses advanced land surface modeling and data assimilation techniques.

Selection of weather data—using weather station, gridded, or reanalysis datasets—for epidemiological analyses involve tradeoffs between the coverage and resolution over space and time and the availability of weather variables. Data selection also depends on the availability of weather station data and researchers' constraints (e.g., topic of study, location and time period, quality and resolution of available outcome data in space and time). If reliable weather data are not available, field studies in environmental health have another option: they can include primary data collection of precipitation measurements using rain gauges or temperature probes as part of the study design, such as the studies in Ecuador,[3, 8] Haiti,[47] India,[25] and Laos.[4] Three studies also supplemented weather station and/or GCD data with local rain gauge and/or temperature probe data.[4, 25, 47] As in the Ecuador study (Carlton et al., 2014),[4] rainfall values from multiple weather monitoring can also be imputed using nonparametric kriging, which provides more accurate climate predictions compared to other interpolation methods such as inverse-distance weighting, nearest-neighbor predictions, and linear regression (Hofstra et al., 2008; Ly et al., 2011; Romero et al., 1998).

Two resources for researchers interested in integrating weather or climate data include (a) the Climate Data Guide by NCAR (https://climatedataguide.ucar.edu/), which presents data summaries, strengths, and limitations on multiple weather and climate datasets, tools, and methods, and (b) a review by Auffhammer and colleagues (2013), which presents a practical discussion of using weather data and climate model output in analyses of the economic impacts of climate change, a comparison of different weather and climate data products, and common mistakes that researchers make and how to avoid them.

Weather station data

Weather stations and satellite data are the sources of most observational weather data (e.g., temperature, precipitation, snow, etc.). Systematic weather monitoring using weather stations began in the 1800s, while weather satellites were first launched in 1960. Station-level weather data for many locations is publicly available for free as raw station data or integrated in databases from institutions like the U.S. National Oceanic and Atmospheric Administration (NOAA). However, using station-level weather data for analyses has a number of limitations due to varying spatial and temporal coverage, especially in areas where weather data collection is not prioritized, as in some regions of low income or low population density. These limitations may hinder analysis, depending on the study location, time period, and weather variables of interest (Auffhammer et al., 2013).

Gridded weather or climate data products

Gridded climate datasets (GCDs) (a.k.a., gridded weather datasets, GWDs) create a balanced panel of weather observations by using spatial extrapolation algorithms to interpolate weather records from weather stations or monitors across space and time over a grid or other fixed spatial scale (Auffhammer et al., 2013). Gridded data products are advantageous because they are often free, easy to import into formats for statistical analyses, and complete in terms of spatial coverage. However, the interpolation process may introduce potential biases over missing observations or areas where there are no weather stations. Ultimately, gridded data products are dependent upon the quality of underlying data. Multiple validation studies have been published recently using at least one GCD for the continental USA (CONUS)—using Daymet (Thornton et al., 1997) and/or PRISM (Daly et al., 1994, 2008; PRISM Climate Group, 2004)—for epidemiology (M. C. Levy et al., 2019; Spangler et al., 2019; N. Thomas et al., 2021) and agriculture (Mourtzinis et al., 2017), and in Switzerland (de Schrijver et al., 2021) using HadUK-grid UKPOC-9 and MeteoSwiss-grid-product for temperature and mortality.

Data assimilation (reanalysis) products

Data assimilation products, referred to as "reanalyses" by the climate community, are an alternative to gridded weather products that also address the limitations of station-level data. Data assimilation products also use observational data, but, unlike traditional gridded weather products, combine them with a physics-based weather model. By combining estimates of weather or climate from the model with observations where data exists and extending to locations where it does not, data assimilation products may be especially advantageous in data sparse regions. Since the underlying model is based on physical laws, global data assimilation products can provide complete spatial coverage of weather data and can be more temporally resolved than gridded weather products. Indeed, reanalysis data may be available the daily or even sub-daily scale. However, it is not possible to force the output of reanalyses to match observational data perfectly because the output has limited resolution, is influenced by the model even when constrained by rich observational data, and reflects the systematic biases and imperfections inherent to modeling. A limited number of reanalysis products are regularly updated because they are difficult and costly to produce (Auffhammer et al., 2013). The ECMRF ERA5 reanalysis product (European Centre for

Medium-Range Weather Forecasts (ECMWF), 2010) was recently validated for temperature and mortality and GLDAS (Rodell et al., 2004) and CHIRPS (Funk 2015) for rotavirus in an 8-site cohort study (Colston et al., 2018).

2.4.5. Meteorological exposure definition

Specific meteorological variables used (TABLE 5 Column L):

The variables used to represent climate and meteorological exposure vary across studies, which complicates interpretation and comparison of results (TABLE 5 Column L). Studies separately estimated models for each meteorological variable or adjusted for more than one exposure within the same model. Most studies included ambient temperature (T) (n=89),[1, 2, 4, 9-45, 47-51, 53-94, 96-98] which usually look at the effects of mean or normalized average temperature, though some rely on maximum and/or minimum temperature or diurnal temperature ranges (DTR) (n=2).[33, 37] Further details on the use of disaggregated measures of temperature (mean, minimum, maximum, etc.) in diarrheal disease studies are available from multiple systematic reviews and meta-analyses (Carlton et al., 2016; Chua et al., 2021, 2022; M. Liang et al., 2021). A somewhat smaller number of studies considered rainfall or precipitation (P or PPT) (n=72),[1- 8, 11-19, 22-26, 29-36, 38, 39, 41, 45-52, 54-57, 59-62, 67-69, 72, 73, 77, 78, 80-82, 84, 86-98] using measures of total (cumulative) or average, extreme or heavy rainfall, or antecedent precipitation characterizing wet, moderate, and/or dry periods, which are further described and evaluated in the systematic review and meta-analysis by Kraay and colleagues (2020). Less than half of the studies accounted for atmospheric water vapor by including exposures for humidity (n=43),[2, 90-11, 13-15, 28, 29, 32-34, 37-39, 41, 44, 47, 60-65, 68, 69, 71, 78, 80, 82-84, 86-92, 94-97] predominantly using measures of relative humidity (RH) (%) instead of specific humidity (SH) (kg/kg). Apparent temperature (AT) (a.k.a. heat index) (n=2),[6, 7] which accounts for the temperature as perceived by humans as a function of temperature and dew point temperature or relative humidity and wind speed, was used as an alternative exposure for temperature in two studies. Many studies that tested precipitation or relative humidity ultimately omitted them from the final models if the variables are not found to be statistically significant in preliminary correlation or regression analyses. Less common meteorological exposures included wind speed (WS) or velocity (m/s) (n=11);[10, 33, 61, 66, 84, 87, 88, 90, 91, 94, 95] atmospheric or air pressure (AP) (n=8)[33, 66, 84, 88, 90, 94, 96, 97] or vapor pressure (VP) (n=2)[66, 95] (kPa, atm, mb, or mmHg); sea surface temperature (SST), mainly in cholera studies; sunshine duration (hours) (n=6);[13, 88, 90, 91, 94, 97] solar radiation (W/m^2) (n=2);[61, 66] and visibility (m or km) (n=1).[10] With the exception of sea surface temperature, all studies incorporating these less common meteorological exposures (WS, AP, VP, sunshine duration, solar radiation, visibility) used multiple meteorological exposures obtained from weather station data and included China as a study setting (n=11).[10, 33, 61, 66, 84, 88, 90, 91, 94, 95, 97] In two recent studies not included in this review, some of these variables (T, P, RH, SH, AP, WS, solar radiation, runoff, soil moisture) were obtained from the reanalysis product GLDAS, as an alternative to weather stations (Colston et al., 2019, 2022).

The relationships between explanatory climate factors and health outcomes were often specified as linear variables, but some studies represented non-linear effects in the relationship between exposure (e.g., temperature, precipitation, relative humidity) and outcome through the use of GAMs,[7, 9, 13, 22, 25, 39, 42, 61, 78, 87, 91, 93] DLNMs,[10, 14, 26, 33, 34, 37, 38, 47, 57, 61, 75, 77, 78, 83, 85, 89, 91, 92] or threshold ("hockey stick") models.[e.g., 74, 81 for temperature] For example, Xu and colleagues (2013) combined a Poisson GLM model with a distributed lag non-linear model (DLNM) examine the effect of diurnal temperature range (DTR) on childhood diarrhea.

Rainfall and precipitation

Exposure definitions for precipitation, which was included in 72 of the 98 studies, typically fall into three categories: absolute measures (e.g., total, average) based on the amount of rainfall in a given time period (e.g., day, week, month), relative measures of heavy or extreme precipitation (e.g., 80th, 90th, 95th percentile), and, least commonly, using indicator variables that reflect relative wet, moderate, and/or dry periods, that aim to test the first flush and dilution theories (Bach et al., 2010; H. Lee et al., 2004; K. Levy et al., 2016).

The first type of precipitation variable is based on absolute measures, typically of the cumulative amount of rainfall in a given period of time. For example, Chen *et al.* (2012) created a categorical precipitation variable with four levels based on different cut points (<130 mm, 130–200 mm, 200–350 mm and >350 mm). Using this kind of categorical variable raises questions about how the cut points are determined, how to decide on a set of cut points if the study area is large and covers multiple climates or micro-climate, and whether a given amount of rain means the same thing in different locations that may vary by climate, soil type, terrain, development, etc.

The second type of precipitation variable is heavy or extreme precipitation. Extreme precipitation is typically a local phenomenon depending on local climatology, which may limit the generalizability of relative measures like percentile indices, particularly in comparative studies or studies occurring over larger spatial areas. The exact definition of the percentile index is also important for inter-study comparisons: not only which percentile is selected to represent heavy rainfall, but also factors such as how to treat days with zero precipitation, the length of the precipitation reference period, and spatial extents involved. To date, AGI-weather studies have not generally adopted the more rigorously defined relative precipitation measures that appear in some climate change literature. These metrics include (a) all-day percentile indices (include dry or zero-precipitation days), (b) wet-day percentile indices (only include wet days over a precipitation threshold and exclude zero-precipitation days, and (c) frequency of exceedance indices (Schär *et al.* (2016). In the climate literature, wet-day indices have been found to be more sensitive to the magnitude and frequency of rainy days and, therefore, the use of all-day indices or exceedance indices are recommended over wet-day indices. However, studies are inconsistent in reporting how their heavy/extreme precipitation indices are defined (e.g., whether or not zero-precipitation days are included, thresholds to define wet days, precipitation cutpoints for percentiles). Furthermore, there is not broad recognition that wet-day and all-day percentiles are not equivalent; for a given location, the precipitation cutpoint defined by a 90th percentile all-day percentile will always be less than an 80th percentile cutpoint. While all-day and wet-day precipitation indices are most common amongst AGI-weather studies, the last of these—exceedance indices—has been used in studies examining relationships between extreme temperature (ETT_{95}), extreme precipitation (EPT_{90}), and health (asthma, salmonellosis, campylobacteriosis) in the state of Maryland (Jiang et al., 2015; Soneja, Jiang, Fisher, et al., 2016; Soneja, Jiang, Upperman, et al., 2016; Upperman et al., 2015) are defined by Equation 1 and Equation 2 (SI Section 2.8.1). The ETT_{95} and EPT_{90} measures are advantageous in that they are designed to account for more resolved weather data matched with outcome data that has a higher level of spatial aggregation. More commonly, AGI-weather studies use all-day or wet-day percentile indexes (e.g., 80th, 90th, 95th, and 99th percentile) to represent exposures from heavy extreme precipitation events.[e.g., 3, 6, 7, 8, 15, 18, 25, 29, 36, 57, 61] Curriero and colleagues (2001) found that precipitation events above the 90th (80th) percentile preceded 51% (68%) of waterborne disease outbreaks in the United States between 1948 to 1994, controlling for season and hydrologic region. One systematic review of AGI-climate studies (K. Levy et al., 2016) found a wide variety of exposure definitions for heavy rainfall and a range of associations: of 10 studies (14 analyses) with quantitative analyses, ten studies reported a significant positive association between AGI and heavy rainfall, three reported a significant negative association, and one found no effect. Similarly, a systematic review of extreme precipitation or temperature and waterborne infections related to drinking water (Guzman Herrador et al., 2015) found 20 studies with positive associations between extreme precipitation and waterborne infections in which 11 of the studies had a significant positive associations, 3 had heterogenous results, and 30% had no association. Most recently, Kraay and colleagues (2020) confirmed the heterogeneity of prior reviews and observed an no statistically significant pooled association between extreme rain and diarrhea (incidence rate ratio IRR = 1.16; 95th CI: 0.946, 1.42) across 13 studies in their meta-analysis (11 of which are included in this review). However, they found effect modification of extreme rain when preceded by a dry period (IRR = 1.16; 95th CI: 0.946, 1.42) and a statistically significant association for studies defining extreme rain based on a storm event (IRR = 2.51, 95% CI: 2.03, 3.10). It is unclear whether variation in the association between extreme precipitation and diarrheal diseases is due to local conditions, modeling choices, and/or exposure definitions.

The third type of precipitation measure is intended to capture wet and dry periods. Developed to test the first flush theory (Bach et al., 2010; Bertrand-Krajewski et al., 1998), they allow comparison of the effects of heavy precipitation following wet or dry periods. These metrics are particularly interesting for climate-health studies because heavy precipitation may lead to different runoff, dilution, and concentration effects as a function of the antecedent weather and their hypothesized mechanisms are discussed in prior studies (Kraay et al., 2020; K. Levy, Hubbard, 2009, Nelson, et al., 2009; Moors et al., 2013). We are aware of eight recent AGI-weather studies that have used antecedent (a.k.a. prior) rainfall to stratify associations with extreme rain (K. F. Bush et al., 2014b; Carlton et al., 2014; Chhetri et al., 2017; Graydon et al., 2022; D. Lee et al., 2019; Mertens et al., 2019; Tornevi et al., 2013, 2015), four of which were included in the systematic review by Kraay and colleagues (2020) and were found to be promising effect modifiers of extreme rainfall (K. F. Bush et al., 2014b; Carlton et al., 2014; Chhetri et al., 2017; Mertens et al., 2019). Together, these studies present five different approaches to defining antecedent precipitation indicator variables, customized to their location(s) and model time step (e.g., daily, weekly) as needed:

1) total of 8-week prior precipitation, defined as tertiles [weekly] (Carlton et al., 2014; D. Lee et al., 2019);
2) X or more dry days (<0.1mm/day) in the prior 60 days, where the threshold of X days

varied by location (Vancouver, Canada: 30d; Hamilton and Toronto, Canada: 35d; Green Bay and Milwaukee, WI, USA: 40d) [weekly] (Chhetri et al., 2017; Graydon et al., 2022);

3) average daily rainfall in the prior 60 days, defined as tertiles [weekly] (Mertens et al., 2019);

4) number of consecutive wet or dry days in the prior 30 days, classified into 6 (dry: >13d, 8-13d, 3-7d; 1-2 wet/dry; wet: 3-7d, >7 wet) or 11 (dry: >5d, 5d, 4d, 3d, 2d, 1d; wet: 1d, 2d, 3d, 4d, >4d) categories [daily] (Tornevi et al., 2013, 2015); and

5) monsoon-based dry (pre-monsoon), moderate (early monsoon), and wet (late monsoon) precipitation seasons [daily] (K. F. Bush et al., 2014b).

Developing, testing, and comparing better, broadly applicable, and, ultimately, more standardized measures of antecedent precipitation exposures for different levels of spatial and temporal (e.g., daily, weekly, month) aggregation and extent is an important new area of research.

Temperature

Temperature exposures are the most commonly included in AGI-weather studies (89 of 98). When used as an exposure or predictor variable in AGI and weather studies, temperature measures can include minimum (T_{min}), mean (T_{mean}), and maximum (T_{max}) temperature, with mean temperature as the most commonly used. Temperature and AGI studies reviewed in a meta-analysis by Carlton and colleagues (2016) are organized by temperature measure (mean, minimum, maximum) and outcome (all-cause, bacterial, viral, protozoan) in TABLE 2. More recently, Chua and colleagues (2021) found that temperature was defined by mean, maximum, and minimum values in their systematic review of 80 temperature and diarrheal disease studies (%): mean (68.8%), minimum (2.5%), maximum (15%), minimum and maximum (10%), and all three values (3.8%). These temperature measures can be aggregated to daily, weekly, and monthly time steps across the different spatial scales and tested for various lags. Alternatively, temperature can be used to define degree days, which are temperature bins used to capture non-linear effects. Degree days are less common in epidemiological studies of AGI and weather (with the exception of the outbreak studies K. M. Thomas et al., 2006; Yang et al., 2012), but have been used in multiple econometrics applications in agriculture (Deschênes & Greenstone, 2007a; Fisher et al., 2012; Schlenker et al., 2006a, 2007a; Schlenker & Roberts, 2006, 2009a). Degree days are more useful for measures defined at longer temporal scales (e.g., monthly or annual rather than daily). Temperature can be included as a linear term, as in a GLM, or smoothed with a spline in a GAM.

TABLE 2. Temperature measure (mean, minimum, maximum) used by 26 quantitative studies of the association between temperature and AGI, reviewed by Carlton *et al.* (2016). Studies are listed by first author and year.

Temperature	All-cause	Bacterial	Viral	Protozoan
Mean (65%)	Checkley 2000 Hashizume 2007 Lama 2004 Onozuka 2010 Singh 2001	Britton 2010 D'Souza 2004 Dewan 2013 Fleury 2006 Kovats 2004 Tam 2006	Atchison 2010 D'Souza 2008 Hashizume, Armstrong, Wagatsuma, et al. 2008 Jagai 2012 Levy 2009 Lopman 2009	
Minimum (12%)	Seidu 2013	Zhang 2010 Ali 2013		
Maximum (23%)	Chou 2010	Bi 2008 Luque Fernández 2009 Traerup 2011 Zhang 2008b		Hu 2007

Source: 26 quantitative temperature and AGI studies from Carlton et al. (2016).

Humidity, apparent temperature, and dew point temperature

Exposure variables that account for water vapor include relative humidity (%), specific humidity (kg/kg), and apparent temperature. Although diarrheal diseases are not airborne, some studies of diarrheal diseases have included relative humidity as an exposure variable or control for confounding (43 of 98), with heterogeneous effects depending on study and model specification. For example, Gleason and Fagliano (2017)[15] controlled for same day temperature and same day relative humidity in their study of the effects of heavy precipitation and drinking water source on gastrointestinal illness in New Jersey. In a study of extreme precipitation ($\geq 90^{th}$ percentile) and AGI hospital admissions in Chennai, India, Bush and colleagues (2014b)[7] accounted for the effects of humidity by including daily average apparent temperature (AT)—a function of air temperature and dew point temperature (Td)—as a potential confounder. If not available with weather data, humidity (relative or absolute) and apparent temperature can be estimated from functions of air temperature (T) and dew point temperature (Td). For futher information, refer to the following resources: humidity best practices (Lawrence et al 2005), summary of relative humidity and absolute humidity equations (Spanger et al, 2019, Supplementary Information); review on humidity and epidemiologic studies, including humidity metrics (Davis et al 2016).

As described in a recent review (X. Wu et al., 2016), disease hosts like mosquitoes, ticks, and fleas react to changes in relative humidity (RH), suggesting mechanisms that may explain why relative humidity has been observed to affect the transmission of vector-borne diseases. Depending on the exposure pathway, relative humidity has also been found to be a relevant predictor of legionellosis (e.g., Fisman et al., 2005), influenza (e.g., Lowen & Steel, 2014), dengue (e.g., Thu et al., 1998), and malaria (Tonnang et al., 2010). Aik and colleagues (2020) found that relative humidity was positively (negatively) associated with diarrheal disease risk with a 1-week (6-week) lag in Singapore and discuss a potential mechanism for relative humidity and diarrheal diseases through food, such that increases in relative humidity may increase the risk of food-borne illness by increasing the viable count of bacterial enteropathogens on contaminated food.

2.4.6. Spatio-temporal scale, resolution, units of analysis (UOA), and aggregation

Study period (TABLE 4 Column D):

The likelihood and severity of diarrheal diseases varies over time, and the epidemiological literature has long acknowledged the need to account for seasonality and interannual variability in diseases (Kelly-Hope & Thomson, 2008). In the studies on weather and diarrhea, most studies span multiple (2+) years (n=95). Study period lengths ranged in length from 2-5 years (n=23), 6-10 years (n=31), 11-15 years (n=19), 16-20 years (n=10), to over 20 years (n=9). Seven studies, including meta-analyses, had multiple study lengths for the same pathogen, varying by study, country, or site (ranging from 1 to 18 years in the studies reviewed).

Study setting: countries (TABLE 4 Column E) and locations (TABLE 4 Column F):

As presented in TABLE 4 Columns E-F and summarized in SI TABLE 9, most studies were conducted in Asia (n=50),[1, 4, 7, 10, 11, 13, 14, 16, 19, 24, 25, 28, 29, 32-36, 39, 45, 46, 48- 51, 56, 61, 63, 65, 69, 72, 73, 78, 82- 98] followed by Africa (n=12),[2, 5, 12, 17, 27, 30, 52, 53, 54, 55, 67, 77] North America (n=10),[6, 15, 18, 22, 40, 42, 44, 47, 57, 58] and Australia (n=9), [37, 38, 41, 43, 64, 71, 76, 80, 81] Europe (n=7), [20, 26, 60, 62, 66, 74, 79] South America (n=4), [3, 8, 9, 21] and Oceania (n=3),[31, 70, 75] in addition to three global studies. [23, 59, 68] Of the countries represented, China has had the highest concentration of studies (n=21),[10, 13, 14, 33, 34, 39, 61, 78, 83-85, 87-92, 94-97] followed by Bangladesh (n=10), [16, 36, 45, 46, 48, 49, 51, 56, 65, 98] and Australia (n=9).[37, 38, 41, 43, 64, 71, 76, 80, 81] Many of the studies in Bangladesh were conducted in Matlab (n=6), which is home to icddr,b and its long legacy of public health and demographic research, especially related to diarrheal diseases and cholera. Studies were conducted in a range of settings based on World Bank country classifications by income: high (n=35), upper-middle (n=30), lower-middle (n=20), low (n=9), and varied income classifications across study countries (n=4) (not reported in table).

Spatial coverage of the analysis (TABLE 4 Column G):

Diarrheal disease outcomes vary significantly across space, though the mechanisms behind this variation are not well understood (Mellor et al., 2016). Some portion of this variation may be attributable to climate, but any attempt to conduct such an analysis is greatly hindered by data limitations in the epidemiological literature related to diarrheal disease outcomes and weather fluctuations. Most of the studies reviewed were conducted in one or a few sites located close together (e.g., hospital, emergency room, or community), with the exception of studies conducted at multiple site locations (n=15)[6, 19, 24, 41, 42, 47, 49, 55, 71, 73, 75, 79, 81, 92, 96] or across multiple countries (n=12) within a single region or across the globe.[10, 20, 23, 31, 46, 53, 59, 60, 62, 67, 68, 74] Some studies were national in coverage (n=13),[11, 12, 17, 26, 31, 34, 35, 47, 55, 69, 70, 93, 97] while others cover smaller sub-national regional,[2, 25, 47, 55, 60, 62, 97] state,[15, 18, 22] provincial,[4, 8, 13, 24, 42, 52, 58, 73, 82, 85, 86, 94] prefecture,[14, 28, 90] or county[44] scales (n=19). Lastly, many studies were conducted in one or more individual cities (n=50), [6, 7, 9, 16, 21, 27, 29, 30, 32, 33, 36-41, 43, 46, 48, 49, 50, 51, 56, 57, 59, 61, 63-65, 68, 71-73, 75-81, 83, 84, 87-89, 91, 92, 95, 96, 98] islands (n=3), [31, 54, 66] communities n=2),[19, 45] or villages (n=1).[3] Some studies analyzed the effects of spatial variation on their results by comparing between countries, sub-national regions, or cities in separate models, by using the smaller areas as the spatial units of analyses (e.g., single model with country-, sub-national, or city-level data), or in meta-analyses or -regressions. The concentration of these multicounty or multi-site studies varied by region: global level (all, Tropics, or low and middle income countries, LMICS) (n=3),[23, 59, 68] East Asia (China, Japan, Taiwan, South Asia) (n=6);[73, 10, 34, 92, 93, 96] Europe (12 countries) (n=5);[97, 26, 60, 62, 74] South Asia (Bangladesh, India + 4 countries) (n=5);[75, 81, 31] Australasia (Australia, New Zealand) (n=4);[47, 41, 71, 75] North America (Canada, Haiti, USA) (n=3);[42, 47, 41] Sub-Saharan Africa (8 countries) (n=3);[49, 67, 24] Central Asia (Kazakstan) (n=1);[97] South Pacific (18 islands) (n=1);[19] Southeast Asia (Cambodia) (n=1).[17] Many of these studies found that the magnitude of climatic sensitivities and the effects of lagged meteorological exposures varied between locations, which results raises the possibility that estimates of climate sensitivity obtained from individual studies may not be widely applicable. Indeed, when working at a broader spatial scale, meta-analytic research often attempts to pool studies such that the meta-analyses sufficiently account for heterogeneity across locations (Gasparrini & Armstrong, 2010).

Model spatial unit of analysis (UOA) (TABLE 4 Column H):

The unit of analysis is the entity a research project seeks insights about; it is the focus and object of study. In statistical studies of relationships between diarrheal disease and climate change or weather variability the unit of analysis may range from the individual to the national population.

Site-specific studies typically focus on individuals or households (n=72),[1, 3-7, 9, 11-16, 19-21, 25, 28-31, 33, 37-41, 43-52, 54-58, 61, 63-66, 69-72, 74-86, 88-92, 95, 96, 98] although a minority use city or community (n=5),[6, 8, 34, 36, 87] larger sub-national regions (e.g., region, province, county, district, metropolitan sub-districts) (n=18),[2, 10, 17, 24, 26, 27, 32, 35, 42, 47, 55, 60, 7-; 73, 87, 93, 94, 97] or country-level (e.g., pooled across a country or island state) (n=4)[31, 53, 60, 62] measures. The regional and global meta-analyses tend to use measures of the prevalence of illness aggregated to the study level as the unit of analysis (n=4).[23, 59, 67, 68]

When considering the model unit of analysis with the available data, it is important for researchers to be mindful of spatial aggregation and spatial boundaries, and their potential effect on the results of the analysis. It is desirable for both the study population and the unit of analysis to be defined at a scale as close as possible to that at which the relationships of interest occur. In environmental health studies, this the aggregation or grouping of data by artificial administrative boundaries and mismatches between the unit of analysis for health outcomes vs. the spatio-temporal resolution of available environmental data present challenges that may affect the accuracy of study outcomes due to issues of ecological or aggregation bias (Shafran-Nathan et al., 2017). In geography and ecology, this unwieldy issue—known as the modifiable areal unit problem (MAUP)—arises with aggregating data into artificial spatial units and grouping them within zones or spatial boundaries (e.g., administrative or ecological) and may lead to variation in results. Often ignored by geographers and public health researchers because it can be difficult to address in practice and can not be ascertained ahead of time (Manley, 2014), MAUP is divided into issues of *scale* (or aggregation), with spatial units of different sizes, and *zone* (or grouping), with different configurations of non-contiguous groups or contiguous zones (Heywood et al., 1988b; Openshaw, 1984c). In epidemiology, MAUP is similar to ecological fallacy, which occurs when results of aggregated data are applied in error to make an inference about an individual in the studied group or zone (Gelman et al., 2001; Openshaw, 1984c, 1984b).

Most human health studies, including those reviewed here, are "organized spatially around human constructs" (Bunch et al., 2011). They use administrative, jurisdictional, or political boundaries (e.g., healthcare facilities, municipalities, postal or zip codes, counties or provinces, states, etc.) as the spatial units of analysis, rather than biogeophysical or socio-ecological units that may be more relevant to understanding and modeling the complex interactions between waterborne illnesses and hydroclimatologyy (Galway et al., 2015; Leyk, Phillips, et al., 2011). Watersheds have been proposed as an alternative unit of analysis as they are watersheds important for water quality protection (e.g., Herrera et al., 2017) and some research has shown that more accurate predictions related to health outcomes resulted from the use of natural boundaries compared to municipal boundaries (Leyk, Phillips, et al., 2011). As discussed by Galway and colleagues (2015), the idea of shifting the unit of analysis for climate-water-health research to the watershed-level is supported by those who argue for combining ecohealth and water resources management to improve human health and well-being (Parkes et al., 2008) and developing the interdisciplinary field of hydroepidemiology (Kay & Falconer, 2008). However, a shift from administrative to watershed boundaries is not without its challenges. Corley and colleagues (2018) explored mapping adverse health outcomes to watershed boundaries and found that hydrologic unit (HU) (e.g., watershed) selection is important: exposures may be combined if the HU is too large, and the population at risk and health outcome incidence may be too small if the selected HU is too small. The authors also discussed the issue of misclassification error when converting the geography of populations to the geography of watersheds and noted that ZIP codes and census blocks overlap better with watersheds than counties. Despite these issues, it would be worthwhile to further explore watersheds as model UOAs and to compare results of AGI-weather models at different spatial scales and with both administrative and ecological boundaries, for example, comparing ZIP code, county, and watersheds of different sizes.

Model time step (TABLE 5 Column I):

Epidemiological studies of the effects of climate on health generally aim to use the finest temporal resolution possible given data constraints. This preference is motivated by concerns over (a) the obscuring of seasonal disease incidence patterns through averaging (measurement error); (b) the inability to control for confounding time-varying factors that may also influence health outcomes (omitted variables bias); and (c) a type of "ecological fallacy" (Koopman & Longini Jr, 1994) in that average or aggregate exposures to the weather variables of interest may be quite different from those that actually drive disease occurrence (model misspecification) (Fisman, 2007). In most cases, the time step is thus daily (n=26),[6, 7, 9, 14, 15, 18, 22, 26, 33, 34, 37-39, 43, 44, 47, 56, 60, 61, 76, 78, 83, 87, 89, 91, 92] weekly or biweekly (n=36)[1, 4, 8, 10, 16, 17, 19, 20, 25, 27, 28, 29, 30, 40, 41, 42, 44, 47, 48, 49, 50, 57, 60, 62, 64, 65, 66, 69, 74, 75, 79, 80, 81, 85, 93, 98] or monthly (n=38),[2, 3, 11, 13, 21, 23, 24, 31, 32, 35, 36, 45, 46, 51, 52, 54, 55, 58, 59, 63, 67, 68, 70, 71, 72, 73, 74, 77, 81, 82, 84, 86, 88, 90, 94, 95, 96, 97] though some studies use different time units for different locations, and a small minority use annual or longer timesteps (n=4)[12, 31, 53, 55] Some studies aggregate daily raw data to a weekly or monthly level in order to ensure sufficient statistical power and case counts in each time bin (Y. Zhang et al., 2010). When a high temporal resolution is not possible due to sparse data on outcomes, several studies may used case-crossover designs (Dixon, 1997; Greer et al., 2008; Maclure & Mittleman, 2000; K. M. Thomas et al., 2006), in which a 'crossover' control is the same individual observed before and after the onset of the outcome.

One global cross-sectional study (Lloyd et al., 2007)[23] unusual in its selection of a coarse temporal resolution. The authors aggregated mean temperature and rainfall to the annual level from diarrheal disease morbidity surveillance studies conducted at intervals of $\leq$2 weeks and noted that associations may be different at different scales; they went on to suggest that insights derived from different time aggregations could thus be applied to different decision-making situations, with long-term aggregations guiding infrastructure development (Lloyd et al., 2007). The authors did not seem to have considered the possibility that associations emerging from annually aggregated data may be unreliable, since an annual time scale may not be an appropriate temporal resolution for seasonal diarrheal diseases and weather, especially when the original diarrheal outcome data had been collected in approximately two-week intervals.

Though the recognition of time aggregation bias is not new (Petersen & Koput, 1992), the modifiable temporal unit problem (MTUP), analogous to the MAUP, is an issue of the temporal dimension that has started to be explored more recently. Temporal effects of MTUP can be defined into issues of temporal aggregation (e.g., daily, weekly, monthly), segmentation (e.g., Sunday, Monday, Tuesday start days for weekly aggregation), and boundary effects (e.g., different study periods and durations) and have been found to have significant effects on the detection of space-time clusters (T. Cheng & Adepeju, 2014). Studies of respiratory disease and dengue, respectively, that have compared different temporal aggregations have highlighted another challenge: the choice of aggregation can obscure connections between disease dynamics and climate risk factors, but the "best" aggregation can change from one risk factor to the next in the same analysis (Gosai et al., 2009; Khormi & Kumar, 2012). When Leyk, McCormick, and Nuckols (2011) compared annual and decadal models of pediatric mortality patterns, rates, and peak timings, they found that aggregating time scales could create spurious relationships that obscured interannual variation. They also reported that variable selection and coefficient values of their models varied with changed temporal aggregation. More recently, Alarcon Falconi and colleagues (2020) explored the effects of temporal aggregation in time series analysis by aggregating diarrheal counts to the daily, weekly, and monthly levels and comparing harmonic regression models of seasonal peak timing and amplitude for three respiratory infections. The authors recommend that researchers conduct sensitivity analyses of the influence of different temporal aggregation units on observed model estimates (Alarcon Falconi et al., 2020).

2.4.7. Non-meteorological covariates (TABLE 5 Column M)

Simultaneous control for a variety of climatic and non-climatic variables is often inconsistent. Control for effect modification by non-climate confounding factors varies widely across studies and there is no standard approach for selection of which specific variables should be included in the models. Covariates or strata included location or spatial indicator variables (e.g., district, county) (n=3);[17, 27, 94] physical variables such as latitude (n=1),[58] elevation (n=1),[32] Normalized Difference Vegetation Index (NDVI) (n=1),[67] dust condition (n=1),[82] stream discharge or flow (n=1),[4] river level (n=6),[1, 16, 32, 48, 65, 98] river temperature (n=1),[44] flood (n=1),[36] El Niño (n=2);[9,21] water quality indicators such as beach closures (n=1)[6] or chlorophyll-a (CHL-a) associated with cholera;[46, 52] aspects of water, sanitation, and/or hygiene (WASH) (n=10);[3, 8, 12, 15, 23, 25, 31, 49, 55, 97] urbanization or remoteness (n=4);[8, 12, 19, 23] socioeconomic status (SES) (n=12),[3, 5, 12, 15, 20, 23, 25, 31, 34, 36, 55, 97] usually a measure of income, wealth, household ownership or assets, education, GDP, or GDP per capita; demographic characteristics (n=15)[3, 12, 15, 18, 23, 25, 26, 33, 34, 35, 36, 37, 55, 56, 93] such as household size,[3, 25] gender,[12, 15, 93] age[12, 15, 18, 23, 25, 26, 33, 35, 56, 93] and race[15]; and other covariates like social cohesion (n=1),[8] breastfeeding status (n=2),[12, 25] disease indicators (e.g., risk, disease presence, frequency or severity; population immunity; outbreak indicator) (n=5),[21, 60, 68, 70, 94] or health service access or characteristics (e.g., hospital, number of beds or technicians) (n=4).[32, 34, 43, 97] All of these factors and more can alter the relationships that may exist between diarrheal disease and climate variables. Indeed, the influence of socioeconomic, geographic, demographic, and environmental confounding is widely recognized due to the many exposure routes, risk factors, and protective WASH interventions (Fewtrell, Kaufmann, Kay, Enanoria, Haller, & Colford Jr., 2005), but available data—which may be obtained from household surveys linked to the outcome data or external datasets—linked to the cases is frequently a limiting factor, particularly for identified data from health facilities.

Hydrological variables

In this review we have generally listed meteorological variables measured in the atmosphere as weather variables and hydrological or hydrometeorological variables measured on land as covariates. Multiple studies included hydrological and hydrometerological factors as covariates (n=8), as many enteric pathogens are transported and transmitted via water. In the reviewed studies, most common hydrological covariate was river level (n=6).[1, 16, 32, 48, 65, 98] Studies also included river temperature (n=1),[44] stream discharge or streamflow (n=1),[4] and flood control status (n=1).[36] Hydrological covariates may be effect modifiers or confounders. For example, streamflow patterns vary in rainfall- versus snowmelt-dominated regimes and streamflow has been found to be effect modifier in British Columbia, Canada (Galway et al., 2015). In Laos, authors hypothesized that stream discharge was negatively associated with diarrhea due to dilution dynamics because higher water tables increased groundwater supply availability (e.g., in wells, municipal water fountains) and contributed to higher discharges. Though hydrological variables have not been used in many weather and diarrheal studies, they are promising potential effect modifiers to explore, subject to data availability, and may provide further insights into mechanisms involved in climate-diarrhea dynamics.

El Niño-Southern Oscillation (ENSO)

The phases of ENSO—including El Niño (warm), La Niña (cold), ENSO Neutral—affect the weather and could therefore be considered as a control variable in the analysis of climate-health relationships, even though ENSO may seem like a meteorological factor at first glance. The El Niño phenomenon was accounted using indicator variables for two in the reviewed diarrheal disease studies (Checkley et al., 2000; Lama et al., 2004),[9, 21] which found increased rates of diarrhea (and diarrhea during periods of cholera[76]) and ambient temperature following El Niño events. ENSO has also been accounted for in studies of salmonellosis (Butler, 2013) and campylobacteriosis (Butler, 2013; Soneja, Jiang, Upperman, et al., 2016). Soneja, Jiang, Upperman, and colleagues (2016) found that the risk of extreme precipitation events increased during La Niña periods, but not El Niño or ENSO Neutral. ENSO has also been associated with cholera (Finger et al., 2014; Hashizume et al., 2011, 2013; Koelle et al., 2005; Pascual, Chaves, et al., 2008; Pascual et al., 2000; Rodó et al., 2002), and is likely to impact other diseases (Kovats et al., 2003). In addition to be included in daily or weekly time series analyses, the monthly extreme heat exposure metric developed by Upperman and colleagues (2015) has been shown to be sensitive to ENSO.

Water, sanitation, and hygiene (WASH)

Despite the long, albeit mixed, history of the burden of disease (Prüss-Ustün et al., 2014; Schmidt, 2014) and impacts on diarrheal diseases from drinking water and sanitation interventions in low- to middle-income countries (Fewtrell, Kaufmann, Kay, Enanoria, Haller, & Colford Jr., 2005; Waddington & Snilstveit, 2009; Wolf et al., 2014), effect modification from water, sanitation, and/or hygiene factors was tested in relatively few of the reviewed studies (n=10).[3, 8, 12, 15, 23, 25, 31, 49, 55, 97] Combinations of WASH-related variables by study included water, sanitation, and hygiene (n=1),[25] water and sanitation (n=5),[3, 12, 23, 31, 55] sanitation and hygiene (n=1),[8] water only (n=2)49,49 and wastewater discharge (n=1).[97] Additional indicators included (un)improved waster and (un)improved sanitation access (n=2),[3, 12] water coverage (%) and sanitation coverage (%) for global studies conducted at the country or study level (n=2),[23, 55] and an extensive list of indicators derived from household and community survey results that included village-level open defecation rate; estimated from rate of reported open defecation from study household; primary water source; presence of water, soap, or towel/cloth at the household hand washing station; indicators for presence of water or a sink at the household hand washing station; indicators for presence of water or flies at the household hand washing station (n=1).[25] It would be valuable to continuing research to expand studies that test effect modification by different WASH indicators relevant to the local context, and to test the validity of using external data sources when survey or individual-level data is not available.

2.4.8. Additional model specification elements

Lag Selection

A possible delayed association between the exposure (e.g., temperature, precipitation, relative humidity) and the outcome can be represented by lagging or shifting the exposure variable behind the outcome by a period of time (the "lag"), thus allowing the modeler to represent the association between the outcome on a given day and exposure on previous days (Bhaskaran et al., 2013). The length of the lag is influenced by the time for pathogen transport in the environment or water distribution system, the incubation period of the specific pathogen after exposure, and the delay before seeking treatment after the onset of illness (Egorov et al., 2003a; Jagai et al., 2015a; Tornevi et al., 2013). Varying by pathogen, the incubation period for viral (R. M. Lee et al., 2013) and bacterial gastroenteritis (Barrett & Fhogartaigh, 2017b; Nataro & Kaper, 1998a) can range from hours to usually less than 7 days, while that of protozoan gastroenteritis is longer: 1-2 weeks (*Giardia lamblia),* 5-28 days (*Cryptosporidium parvum)*, and 1-4 months (*Entamoeba histolytica)* (Marshall et al., 1997).

Common approaches to represent the delay between exposure and health outcome include single-day lag terms modeled one at a time; lagged multi-day moving averages[7] and distributed lag models (DLMs) (Schwartz, 2000b; Zanobetti et al., 2000), in which lag terms are modeled together and may be either unconstrained or constrained to reduce collinearity (Bhaskaran et al., 2013; Gasparrini et al., 2010). Constrained DLMs may be lag-stratified (Armstrong, 2006) or apply more complex constraints such as polynomial or other smoothing functions of lag time (Schwartz, Spix, Touloumi, Bachárová, Barumamdzadeh, Le Tertre, et al., 1996). Distributed non-linear lag models

[7] A model that defines a predictor as "the moving average of exposures in the previous *L* days" is a special case of a distributed lag model (DLM) (Gasparrini et al., 2010).

(DLNMs) (Gasparrini, 2011, 2014; Gasparrini et al., 2010, 2017; Gasparrini & Leone, 2014) expand on DLMs to include non-linear exposure-response relationships. Lag structure may also take a uniform-weighted (Jagai et al., 2015c), gamma distribution (J. N. S. Eisenberg et al., 1998), or Poisson distribution (Egorov et al., 2003a). Although Jagai and colleagues (2015c) used a uniform-weighted lag structure, they recommended that estimates using a Poisson or gamma distribution may be accurate. Amongst the reviewed studies, unconstrained and constrained DLM (n=13)[7, 16, 17, 18, 28, 29, 46, 48, 49, 62, 69, 70, 74] and DLNM (n=18)[10, 14, 26, 33, 34, 37, 38, 47, 57, 61, 75, 77, 78, 83, 85, 89, 91, 92] lag structures were used (TABLE 5 Column J). Authors are increasingly accounting for lagged exposure-response non-linearities; DLNMs have been employed in more recent studies (89% published in or after 2014) compared to DLMs (43%).

Though not specifically reported in this review, many studies included lags of climatic exposure variables (e.g., temperature, precipitation, humidity, etc.); the details of the reported and/or max lags included in the models are available in the original systematic reviews (Chua et al., 2021, 2022; Kraay et al., 2020; M. Liang et al., 2021; P. Wang et al., 2021). Prior literature provides further information on methods to incorporate lags into climate-health time series studies (Bhaskaran et al., 2013), lag selection techniques (Peng, Dominici and Louis, 2006), and lagged associations in climate-health studies for 14 causes of morbidity and mortality (including 20 diarrheal, 8 cholera, and 12 salmonella studies) is available in a recent systematic review (P. Wang et al., 2021). Wang and colleagues (2021) reported on four components related to lags (lag design, maximum lag, reported lag, lag selection criteria) and found lagged predictors varied by outcome: diarrhea (temperature, rainfall, humidity, flood, atmospheric pressure), salmonellosis (temperature, rainfall, humidity), and cholera (temperature, rainfall), with temperature and rainfall as the most common. When aggregated by country across the three diarrheal outcomes, ambient temperature (mean, minimum, and maximum) was associated with diarrhea for an average of 3-6 weeks and rainfall for 2-10 weeks in most countries (P. Wang et al., 2021), though shorter lags 1 week (e.g., Drayna et al., 2010) or less have also been used.

Controls for seasonality and long-term (secular) trends and other time confounders

Effective modeling of the association between exposure sand outcomes in time series analysis relies on controlling for time-variant variables. Time-variant confounders of the relationship between weather and diarrheal diseases may include long-term trends, seasonal patterns (seasonality), within-month variation, weekend or day-of-the-week (DOW), school vacations and holidays, and residual autocorrelation, which may stem from commercial, healthcare access, demographic, or epidemiological factors similar to those described for the relationship between turbidity and acute gastroenteritis (Beaudeau, Le Tertre, et al., 2012). The majority of studies controlled for long-term trends (n=62)[6, 7, 10, 11, 13, 14, 16-20, 22, 24, 26, 28, 29, 31, 33, 34, 37-44, 48, 50, 55, 57, 60-62, 64, 65, 68, 69, 71-81, 83-85, 87-94, 96, 98] and month, season, or seasonality (n=65).[6, 7, 9-20, 22, 24, 26, 28-34, 37-42, 44-46, 48, 50, 54, 57, 60-62, 64, 65, 69, 71-75, 77, 78, 80-83, 85-87, 89-93, 95, 96, 98] Fewer studies also included terms for autocorrelation of the outcome variable, also known as autoregressive term(s) (e.g., Y_{-1}) (n=41),[4, 8, 9, 11, 16, 20, 25, 27, 28, 30, 34, 40-43, 45, 48, 50, 51, 53, 54, 59-61, 63, 64, 68-70, 72-74, 76, 80, 81, 84, 85, 90, 92, 95, 96] holidays (n=20),[16, 20, 22, 26, 38, 40, 42, 43, 48, 57, 60-62, 65, 69, 74, 76, 78, 87, 98] and weekend[60] or day-of-week[6, 7, 14, 22, 26, 33, 38, 39, 43, 61, 76, 78, 83, 87, 89] (n=16). Methods to control for long-term trends and seasonality reflected the three common strategies to control for seasonality and long-term trends in time series models, summarized in TABLE 3: time stratified-models, periodic functions, and flexible spline functions (Bhaskaran et al., 2013). In the first strategy, time-stratified models use a single indicator variable per time interval (e.g., calendar month, year) to control for seasonality and long-term trends respectively. The second strategy uses pairs of sine and cosine functions, also known as Fourier terms, harmonics, or harmonic regression, to capture regular seasonal patterns.[e.g., 8, 20, 23, 42, 50, 65, 67, 76] To capture long-term non-seasonal trends, these functions need to be coupled with a separate control function for calendar time. The third strategy uses flexible spline functions to control for seasonality and long-term trends by fitting multiple polynomial (e.g., cubic or natural spline) curves and smoothly joining them together to span the full study period.[e.g. 4, 33] Spline functions are created by generating and including basis variables of the principal time variable in the time series model. Each join, known as a "knot," represents a degree of freedom, and the number of knots determines the flexibility of the spline function. Seven knots per year[8] is a common choice to balance seasonality, time trend control, and sufficient information to estimate exposure effects (Dominici, Zeger, et al., 2000). A fourth (non-time series) strategy, not discussed in Bhaskaran *et al.* (2013), uses (seasonal) autoregressive integrated moving average (with exogenous variables), or (S)ARIMA(X) models (n=6)[45, 54, 80, 84, 95, 96] instead of a standard time series regression model. Examples of seasonal and long-term trend control terms from five selected AGI-weather studies are listed in SI TABLE 12.

TABLE 3. Advantages and limitations of 4 strategies for controlling for seasonal and long-term trends (adapted from Bhaskaran et al., 2013; Kane et al., 2014; Valipour, 2015).

Control Strategy	Pros	Cons
(1) Time-stratified model[a]	• easy to understand • captures main long-term patterns	• may use many model parameters • "implicitly assumes biologically implausible jumps in risk between adjacent time intervals" (Bhaskaran et al., 2013)
(2) Periodic functions[a]	• smoothly models long-term patterns • uses relatively few model parameters	• more mathematically complex than (1) • modeled seasonal pattern is forced to be uniform across years, therefore may not capture natural data pattern • cannot capture long-term non-seasonal trends (*solution*: add function(s) of calendar time)
(3) Flexible spline functions[a]	• smoothly models long-term patterns • modeled seasonal patterns are allowed to vary between years • captures long-term non-seasonal trends	• more mathematically complex than (1) or (2) (*solution*: major statistical packages provide functions to generate spline basis variables)
(4) SARIMA(X)[b,c]	• account for non-stationary processes by incorporating differences to remove trend and seasonal components [b] • model selection can be automated to maximize prediction accuracy [b] • accounts for dynamic relationships over time [b]	• relationships between independent and dependent variables assumed linear [b] • s.d. of errors over time assumed constant [b] • SARIMA models are sensitive to the periodic term s [c]
References : [a]Bhaskaran *et al.* (2013) ; [b]Kane *et al.* (2014) ; [c]Valipour (2015)		

[8] $n_{knots} = (n_{calendar\ years} \cdot 7) - 1$ for daily health (e.g., mortality) data (Bhaskaran et al., 2013)

All are applicable to a variety of time series methods and must be adapted for the time unit of observation (e.g., day, week, month, etc.).

Seasonal sub-analyses

Many studies use seasonal sub-analysis, which defines season as a fixed effect variable with (or without) an interaction term or by strata, to understand the effect modification of (or control for) season on the relationship between diarrheal disease and weather (Curriero et al., 2001; Gleason & Fagliano, 2017; Jagai et al., 2015; D. Lee et al., 2019; Nichols et al., 2009). As previously discussed, viral AGI occurs more often in the cooler months, while bacterial AGI occurs more often in warmer months. For studies using all-cause AGI as one of the outcomes, stratifying models by season can be a useful method to investigate weather patterns that may vary due to changes in the etiology underlying AGI throughout the year. In temperate climates, four seasons may be defined as 3-month meteorological seasons, with spring starting in March/September for the Northern/Southern Hemispheres, (Ahmed et al., 2013; Curriero et al., 2001b; Jagai et al., 2015; Nichols et al., 2009; Upperman et al., 2015) or based on annual solstice and equinox dates (D. Lee et al., 2019). Some researchers have defined seasons as hot and cold (Gleason & Fagliano, 2017; Sugg et al., 2016) or based on monsoon patterns (pre-monsoon, early monsoon, late monsoon) (K. F. Bush et al., 2014b). In non-temperate or tropical climates, seasons have been more commonly defined based on rain patterns (e.g., rainy, dry) (Kraay et al., 2020).

Public holidays

Controlling for national and/or sub-national holidays is a common practice in time series analysis (n=20).[16, 20, 22, 26, 38, 40, 42, 43, 48, 57, 60-62, 65, 69, 74, 76, 78, 87, 98] Holidays may be associated with AGI because people may be exposed to contaminated food during holiday parties or gatherings. Furthermore, the occurrence of a holiday may be confounding if a sick person chooses to delay their visit to a health facility until after a holiday, or is obliged to wait because the health facility is closed.

To control for public holiday effects in daily time series analyses of AGI and weather, indicator variables may be created for the holiday day itself or for the period including the day before and the day after the holiday in a daily time series model (Tornevi et al., 2015) or, in a weekly time series, an indicator variable could designate a week that contains a public holiday (Chhetri et al., 2017).

2.4.9. Further considerations

Considering the role of adaption

There is inevitably great uncertainty regarding the trajectory and long-term dynamics of climate change, and this uncertainty compounds when considering its potential impacts on climate-sensitive diseases (Kolstad & Johansson, 2011; World Health Organization (WHO), 2003).The question of adaptation further complicates matters. WASH factors are known to alter the risks of diarrheal disease, especially child diarrhea and stunting in children (Fewtrell, Kaufmann, Kay, Enanoria, Haller, & Colford, 2005b; Fink et al., 2011). Thus, societies could invest in better WASH infrastructure in order to decrease vulnerability to diarrheal disease outbreaks arising from climate change (World Health Organization (WHO), 2003). The role of such adaptation has been a matter of some debate in the larger and more established literature on climate and agriculture. Statistical models that assess the effects of short-term variability or extreme events on crop yields, for example, may underestimate the potential for adaptation to gradual changes (Haines & Patz, 2004; McMichael et al., 2006). On the other hand, low-resolution models that exploit cross-sectional variation alone in order to capture adaptation to "local conditions" at coarse scales may suffer from omitted variables (Rosenthal, 2009; Schlenker et al., 2005).

Omitted variables seem a particular threat to the modeling of health effects of climate change. The epidemiological literature has long understood the importance of clarifying causality in the context of effect modification by factors such as population density, technology and the quality of health systems, public health infrastructure, standard of living, local environmental conditions, and individual behavioral factors. Such control variables or effect modifiers are conspicuously absent from many of the statistical models linking climate and health, as discussed in 2.4.7, but are increasingly present in more recent studies.

2.5. Conclusions

As witnessed in this review, the range of practices in the literature is vast, which poses a number of critical challenges. Absence of common methodological guidelines for climate-health statistical analyses combined with differences in specific health outcomes (all-cause vs. pathogen-specific), health measures (relative risk, odds ratios, model coefficients), weather variable definitions, and control variables or effect modifiers, continues to make comparisons between results difficult, a persistent finding noted by other reviewers over the past decade (Butler & Hall, 2009; Guzman Herrador et al., 2015; Kraay et al., 2020; K. Levy et al., 2016). It is hard to determine to what extent the variability between results is due to differences between models or heterogeneity across study sites.

Metrics of precipitation and extreme precipitation are particularly challenging, since options abound and there is as yet no clear mechanistic justification for choosing between them in any given context. Explicitly documented definitions (e.g., all-day vs. wet-day) and cutpoints for heavy or extreme precipitation should facilitate inter-study comparison and meta-analysis. There is still much to be investigated about the role of the concentration-dilution hypothesis in climate-health relationships, and we echo Kraay and colleagues (2020) in calling for further research into measures of antecedent precipitation and their use to test effect modification of the association with extreme precipitation.

Spatiotemporal aggregation presents another challenge, and another research opportunity. Although there were studies at all levels of spatial coverage represented amongst all temporal time steps (daily, weekly or biweekly, monthly, annual or multiple cross-sections), the majority of analyses conducted in a single city or community were conducted at the daily scale (39%), the majority of analyses across a sub-national region (50%) were at the weekly scale, and the majority of multi-site (43%), country (43%), or multi-country (42%) analyses were at the monthly scale. Many of the studies that featured the highest spatial coverage were less likely to have higher degrees of temporal variation. This tradeoff simplifies data collection and limiting one of the two dimensions is further appealing because it strengthens the internal validity of the study by reducing the threat of confounding by unobserved (time-varying or spatially heterogeneous) disease dynamics. However, if the climate-health relationship is context-dependent in ways that are not fully understood by the researcher, the spatiotemporal simplification may obscure results and be unsuitable for extrapolation at scale.

The meta-regression approach, which has only recently been tapped in the epidemiological literature on climate change, offers an alternative to ever-more-complex study designs when the goal is to incorporate findings from epidemiological studies into climate change impact projections. Meta-analyses strive to encompass—and in the best case are able to explain—the heterogeneity in relationships between climate and health across both time and space, which may improve generalizability, particularly if some of the variation across locations can be explained. However, limitations on the number of explanatory parameters force trade-offs between model efficiency and site-specific information (Gasparrini & Armstrong, 2010), and the meta-regression and meta-analysis approaches remains prone to biases related to the many aforementioned shortcomings in empirical assessments of disease sensitivity to weather variables.

These challenges are not unrelated to the lack of fundamental understanding of the drivers of relationships between climate and different health outcomes, which likely vary across diseases and locations in complicated ways. Statistical approaches are not well suited for uncovering these relationships and the need for systems-based or mechanistic approaches has been expressed (Mellor et al., 2016). In the absence of a more theoretical and mechanistic understanding of disease dynamics, well-designed statistical modeling studies at local or regional scale incorporating both effect modification important risk factors and sensitivity analysis of outcomes, exposures, and spatio-temporal aggregation can help identify climate-sensitive risk factors in the climate-health relationships.

2.6. Tables

TABLE 4. (Part I: A-H) Review of studies of the effects of climate on diarrheal disease outcomes in the epidemiological literature, highlighting eight of thirteen methodological issues discussed, arranged by diarrheal or gastroenteritis etiology (all-cause, campylobacteriosis, cholera, cryptosporidiosis and giardiasis, *E. coli* enteritis, norovirus, rotavirus, salmonellosis, shigellosis/bacillary dysentery (bac. dys.), typhoid fever).

#	First author (Year)	A Etiology (outcome measure)[1]	B Ages (years)	C Outcome data source[2]	D Study period (years)	E Country[3]	F Location(s)	G Spatial coverage	H Model UOA[4]
1	Alexander (2018)	all-cause (cases)	<5	surveillance (health system based)	2007-2017 (11)	Botswana	Chobe district	district	individual
2	Azage (2017)	all-cause (cases)	<5	surveillance (health system based)	2013-2015 (3)	Ethiopia	Amhara Region (Awi; West and East Gojjam Zones)	sub-national region	district
3	Bhavnani (2014)	all-cause (15-day prevalence)	all	community	2008-2009 (0.5)	Ecuador	Borbón	village	household
4	Boithias (2016)	all-cause (incidence)	all	surveillance (health system based)	2010-2012 (2)	Laos	Luang Prabang Province	sub-province (5 districts)	individual
5	Busch (2019)	all-cause (cases)	≥18	community	2013-2014 (2)	Uganda	Kanungu district	district	household
6	Bush (2014a IJERPH)	all-cause (cases)	≥65	hospital	2000-2006 (7)	USA	12 cities (Buffalo, NY; Chicago, IL; Cleveland, OH; Detroit, MI; Erie, PA; Gary, Indiana; Grand Rapids, MI; Milwaukee, WI; Minneapolis, MN; Rochester, NY; Rockford, IL; Toledo, OH)	multi-site (city)	individual; city
7	Bush (2014b EHP)	all-cause (cases)	all	hospital	2004-2007 (4)	India	Chennai, Tamil Nadu	city	individual
8	Carlton (2014)	all-cause (incidence)	all	community	2004-2007 (3)	Ecuador	Esmeraldas province (19 villages)	province	community (village)
9	Checkley (2000)	all-cause (cases)	<10	hospital	1993-1998 (6)	Peru	Lima	city	individual
10	Chen (2018)	all-cause (incidence)	all	surveillance (health system based)	2012-2016 (5)	China; Taiwan; Japan	Hong Kong SAR; Taiwan (3 regions), Japan (8 regions)	multicountry	region
11	Chou (2010a)	all-cause (incidence)	all; <15; 15-39; 40-64	surveillance (health system based)	1996-2007 (12)	Taiwan	nationwide	country	individual
12	Epstein (2020)	all-cause (cases)	≤2	community	2009-2012 (4)	Uganda	nationwide	country	household
13	Fang (2019)	all-cause (prevalence)	<6, >20	surveillance (health system based)	2013-2017 (5)	China	Jiangsu	province	individual
14	Gao (2020)	all-cause (prevalence)	all	surveillance (health system based)	2013-2018 (6)	China	Wuxi	prefecture	individual
15	Gleason (2017)	all-cause (cases)	all; <6, 6-64, ≥65	surveillance (health system based)	2009-2013 (5)	USA	New Jersey	state	individual
16	Hashizume (2007)	all-cause (incidence)	all; <15; 15-29; >29	surveillance (health system based)	1996-2002 (7)	Bangladesh	Dhaka	city	individual

#	First author (Year)	A Etiology (outcome measure)[1]	B Ages (years)	C Outcome data source[2]	D Study period (years)	E Country[3]	F Location(s)	G Spatial coverage	H Model UOA[4]
17	Horn (2018)	all-cause (cases)	all	surveillance (health system based)	1997-2014 (18)	Mozambique	nationwide	country	administrative district; region
18	Jagai (2015)	all-cause (incidence)	<5; 5-19; 20-64; ≥65	hospital	2003-2007 (5)	USA	Massachusetts	state	region
19	Kulinkina (2016)	all-cause (cases)	all (child <5 in HH)	community	2010-2012 (2)	India	Tamil Nadu (sites: 2 urban, 3 rural)	multi-site (community)	individual
20	Lake (2009)	all-cause (incidence)	all	surveillance (health system based)	1974-2006 (30)	England & Wales	England & Wales	multicountry (pooled)	individual
21	Lama (2004)	all-cause (cases)	>13	hospital	1991-1998 (8)	Peru	Lima	city	individual
22	Lin (2016)	all-cause (cases)	all	hospital	1991-2004 (14)	USA	New York	state	region
23	Lloyd (2007)	all-cause (incidence)	<5 (0-5 mo, 6-11 mo, 1, 2, 3, 4)	community surveillance (published articles)	1954-2000 (≥1 by study)	Global (LMICS)	WHO regions or country (by study)	WHO global region; multicountry (study)	study
24	McIver (2016)	all-cause (cases)	all	surveillance (health system based)	1997-2012 (16)	Cambodia	provinces (11 of 24)	multi-site (province)	province
25	Mertens (2019)	all-cause (7-day prevalence)	<6	community	2008-2009 (>1)	India	Tiruchirappalli, Tamil Nadu (25 villages)	sub-national region	household
26	Morral-Puigmal (2018)	all-cause (incidence)	all; <2; 2-15; 16-64; ≥65	surveillance (health system based)	1997-2013 (17)	Spain	nationwide	country	province
27	Musengimana (2016)	all-cause (incidence)	<5	surveillance (health system based)	2012-2014 (2)	South Africa	Cape Town (8 sub-districts)	city	metro sub-districts (8)
28	Onozuka (2010)	all-cause (cases)	all	hospital	1999-2007 (9)	Japan	Fukuoka prefecture	prefecture	individual
29	Phung (2015)	all-cause (cases)	all	surveillance (health system based)	2004-2011 (8)	Vietnam	Can Tho	city	individual
30	Seidu (2013)	all-cause (cases)	all	community	2008-2009 (1)	Ghana	Tamale	city	individual
31	Singh (2001)	all-cause (cases)	all; <1	surveillance (health system based)	1986-1994 (9)	Pacific island countries (18)	American Samoa; Cook Islands; Fiji; French Polynesia; Guam; Kiribati; Marshall Islands; Nauru; New Caledonia; Niue; Palau; Samoa; Solomons; Tokelau; Tonga; Tuvalu; Vanuatu; Wallis	multicountry (islands)	island
31	Singh (2001)	all-cause (cases)	all; <1	surveillance (health system based)	1978-1989 (12)	Fiji	nationwide	country	individual
32	Thompson (2015)	all-cause (incidence)	<16	hospital	2005-2010 (2; 6 by location)	Vietnam	Ho Chi Minh City (24 districts)	city	district
33	H. Wang (2019)	all-cause (incidence)	all; 0-1, 2-9, 10-29, ≥30	surveillance (health system based)	2006-2017 (12)	China	Guangzhou	city	individual
34	P. Wang (2021)	all-cause (incidence)	all	surveillance (health system based)	2014-2016 (3)	China	nationwide (270 cities)	country (city)	city

#	First author (Year)	A Etiology (outcome measure)[1]	B Ages (years)	C Outcome data source[2]	D Study period (years)	E Country[3]	F Location(s)	G Spatial coverage	H Model UOA[4]
35	Wangdi & Clements (2017)	all-cause (incidence)	all; <5, ≥5	surveillance (health system based)	2003-2013 (11)	Bhutan	nationwide (districts)	country	district
36	Wu (2014)	all-cause (prevalence)	<5	community	2000-2006 (7)	Bangladesh	Matlab	city	community (bari)
37	Xu (2013)	all-cause (cases)	<5	hospital	2003-2009 (7)	Australia	Brisbane	city	individual
38	Xu (2014)	all-cause (cases)	<1; 1; 2-4; 5-14; <15	hospital	2001-2010 (10)	Australia	Brisbane	city	individual
39	Zhou (2013)	all-cause (cases)	all	hospital	2008-2010 (3)	China	Shanghai	city	individual
40	Allard (2011)	campylobacteriosis (incidence)	all	surveillance (health system based)	1990-2006 (17)	Canada	Montreal	city	individual
41	Bi (2008)	campylobacteriosis (cases)	all	surveillance (health system based)	1990-2005 (16)	Australia	Adelaide & Brisbane	multi-site (city)	individual
42	Fleury (2006)	campylobacteriosis (cases)	all	surveillance (health system based)	1992-2000 (9)	Canada	Alberta; Newfoundland-Labrador	multi-site (province)	sub-national regions (provinces)
20	Lake (2009)	campylobacteriosis (incidence)	all	surveillance (health system based)	1989-2006 (18)	England & Wales	England & Wales	multicountry (pooled)	individual
43	Milazzo (2017)	campylobacteriosis (incidence)	all	surveillance (health system based)	1990-2012 (23)	Australia	Adelaide	city	individual
44	White (2009)	campylobacteriosis (incidence)	all	surveillance (health system based)	1994-2007 (14)	USA	Philadelphia County, PA	county	individual
45	Ali (2013)	cholera (incidence)	all	surveillance (hospital based)	1988-2001 (14)	Bangladesh	Matlab	community (upazila)	individual
46	de Magny (2008)	cholera (cases)	all	surveillance (hospital based)	1998-2006 (9)	Bangladesh; India	Matlab, Bangladesh; Kolkata, India	multicountry (city)	individual
47	Eisenberg (2013)	cholera (incidence)	all	hospital; IDP camps; national health dataset	2010-2011 (<1)	Haiti	nationwide	country; multi-site (sub-national regions)	sub-national regions; individual
48	Hashizume (2008a Epi)	cholera (cases)	all	surveillance (health system based)	1996-2002 (8)	Bangladesh	Dhaka	city	individual
49	Huq (2005)	cholera (incidence)	all	surveillance (hospital based)	1997-2000 (4)	Bangladesh	Bakerganj; Chhatak; Chaugachha; Matlab	multi-site (city)	individual
50	Luque Fernández (2009)	cholera (cases)	all	outbreak surveillance (NGO medical registries)	2003-2006 (4)	Zambia	Lusaka	city	individual
51	Matsuda (2008)	cholera (cases)	<10	hospital	1983-2002 (20)	Bangladesh	Dhaka	city	individual
52	Mendelsohn (2008)	cholera (incidence)	all	epidemiological case reports (WHO)	2000-2001 (2)	South Africa	KwaZulu-Natal province	province	individual

#	First author (Year)	A Etiology (outcome measure)[1]	B Ages (years)	C Outcome data source[2]	D Study period (years)	E Country[3]	F Location(s)	G Spatial coverage	H Model UOA[4]
53	Paz (2009)	cholera (cases)	all	global health records (WHO)	1971-2006 (36)	SE Africa (8 countries)	Uganda; Kenya; Rwanda; Burundi; Tanzania; Malawi; Zambia; Mozambique	multicountry (sub-continent region)	country
54	Reyburn (2011)	cholera (cases)	all	surveillance (health system based)	2002-2008 (7)	Tanzania	Unguja, Zanzibar	island	individual
55	Trærup (2011)	cholera (cases & fatalities)	all	national health datasets	1998-2004 (7)	Tanzania	21 regions	multi-site (sub-national regions)	sub-national regions
55	Trærup (2011)	cholera (cases)	all	national health datasets	1998-2004 (7)	Tanzania	nationwide	country	individual
55	Trærup (2011)	cholera (cases & fatalities)	all	national health datasets	1977-2004 (28)	Tanzania	nationwide	country	individual
56	Wu (2018)	cholera (cases)	all; <18, 18-64, ≥65	surveillance (hospital based)	1983-2009 (26)	Bangladesh	Matlab	city	individual
57	Chhetri (2017)	cryptosporidiosis and giardiasis (cases)	all	surveillance (health system based)	1997-2009 (13)	Canada	Vancouver, British Columbia	city	individual
58	Bifolchi (2014)	*E. coli* enteritis (incidence)	all	hospital	2004-2011 (8)	Canada	Alberta	province	individual
42	Fleury (2006)	*E. coli* enteritis (cases)	all	surveillance (health system based)	1992-2000 (9)	Canada	Alberta; Newfoundland-Labrador	multi-site (province)	sub-national regions (provinces)
59	Philipsborn (2016)	*E. coli* enteritis (incidence)	all	hospital & community (published articles)	1974-2004 (1-5 by study)	Global	28 studies	multicountry (city)	study
60	Lopman (2009)	norovirus (incidence)	all	surveillance (health system based)	1993-2006 (14)	England & Wales	England (East Midlands; Yorkshire & Humberside; Northwest England; West Midlands; Southeast; East of England; Southwest; London; Northeast England); Wales	multi-site (regions)	country; region
61	Wang (2018b STOTEN)	norovirus (cases)	<5	hospital	2002-2011 (10)	China	Hong Kong SAR	city	individual
62	Atchison (2010)	rotavirus (incidence)	<5	surveillance (health system based)	1993-2007 (15)	England, Wales, Scotland, the Netherlands	England (Northeast; Northwest; Yorkshire & Humberside; East Midlands; West Midlands; East of England; London; Southeast; Southwest); Wales; Scotland; Netherlands	multicountry (countries & sub-national regions)	sub-national regions (England); country (Wales, Scotland, The Netherlands)
63	Celik (2015)	rotavirus (proportion; cases)	<5	hospital	2006-2012 (7)	Turkey	Sivas City	city	individual
64	D'Souza (2008)	rotavirus (cases)	<5	hospital	1993-2003 (11)	Australia	Brisbane; Canberra; Melbourne	city	individual
65	Hashizume (2008b E&I)	rotavirus (cases)	all	surveillance (health system based)	1996-2001 (6)	Bangladesh	Dhaka	city	individual

#	First author (Year)	A Etiology (outcome measure)[1]	B Ages (years)	C Outcome data source[2]	D Study period (years)	E Country[3]	F Location(s)	G Spatial coverage	H Model UOA[4]
66	Hervás (2014)	rotavirus (cases)	<5	surveillance (health system based)	2000-2010 (11)	Spain	Mallorca	island	individual
67	Jagai, Sarkar, et al. (2012)	rotavirus (z-score)	all	hospital & community (published articles)	1966-2010 (1-12 by study)	South Asia (5 countries)	Bangladesh; India; Nepal; Pakistan; Sri Lanka	multicountry (study site)	study
68	Levy (2009)	rotavirus (incidence)	all; infants & young children; <2; <3; <5; <8; <12	hospital (published articles)	1975-2003 (1-10 by study)	Tropics (13-15 countries)	Dhaka, Bangladesh; Matlab, Bangladesh; Hong Kong; Jeddah, Saudi Arabia; Mexico City, Mexico; Yangon, Myanmar; Coro, Venezuela; Caracas, Venezuela; San Jose, Costa Rica; Addis Ababa, Ethiopia; Guayaquil, Ecuador; Yogyakarta, Indonesia; Apia, Western Samoa; Lusaka, Zambia; Minas Gerais, Brazil; Rio de Janeiro, Brazil; 15 countries	multicountry (city)	study
61	Wang (2018b STOTEN)	rotavirus (cases)	<5	hospital	2002-2011 (10)	China	Hong Kong SAR	city	individual
69	Aik (2018)	salmonellosis (incidence)	all	surveillance (health system based)	2005-2015 (11)	Singapore	nationwide	country	individual
70	Britton (2010)	salmonellosis (cases)	all	surveillance (health system based)	1965-2006 (42)	New Zealand	nationwide	country	individual
71	D'Souza (2004)	salmonellosis (incidence)	all	surveillance (health system based)	1991-2001 (11)	Australia	Adelaide; Brisbane; Melbourne; Perth; Sydney	multi-site (city)	individual
42	Fleury (2006)	salmonellosis (cases)	all	surveillance (health system based)	1992-2000 (9)	Canada	Alberta; Newfoundland-Labrador	multi-site (province)	sub-national regions (provinces)
72	Grjibovski (2013)	salmonellosis (cases)	all	surveillance (health system based)	1992-2008 (17)	Russia	Arkhangelsk City	city	individual
73	Grjibovski (2014)	salmonellosis (cases)	all	surveillance (health system based)	2000-2010 (11)	Kazakhstan	Astana; Almaty; North Kazakhstan; South Kazakhstan (4 regions)	multi-site (city or province)	administrative unit
74	Kovats (2004)	salmonellosis (cases)	all	surveillance (health system based)	1984-2002 (3-18 by country)	Europe (11 countries)	Poland; Scotland; Denmark; England & Wales; Estonia; Netherlands; Czech Republic; Switzerland; Slovak Republic; Spain	multicountry	individual
20	Lake (2009)	salmonellosis (incidence)	all	surveillance (health system based)	1981-2006 (26)	England & Wales	England & Wales	multicountry (pooled)	individual
75	Lal (2016)	salmonellosis (incidence)	all	surveillance (health system based)	1997-2007 (11)	New Zealand	Auckland; Christchurch; Wellington	multi-site (city)	individual
76	Milazzo (2016)	salmonellosis (incidence)	all	surveillance (health system based)	1990-2012 (23)	Australia	Adelaide	city	individual

#	First author (Year)	A Etiology (outcome measure)[1]	B Ages (years)	C Outcome data source[2]	D Study period (years)	E Country[3]	F Location(s)	G Spatial coverage	H Model UOA[4]
77	Thindwa (2019)	salmonellosis (incidence)	all	hospital	2000-2010 (10)	Malawi	Blantyre	city	individual
78	Wang (2018a EI)	salmonellosis (incidence)	all	hospital	2002-2011 (10)	China	Hong Kong SAR	city	individual
79	Yun (2016)	salmonellosis (incidence)	all	government reports	2001-2004 (4)	Germany	Berlin; Munich	multi-site (city)	individual
80	Zhang (2008a IJB)	salmonellosis (cases)	all	surveillance (health system based)	1990-2004 (forecast: 2004) (15)	Australia	Adelaide	city	individual
81	Zhang (2010)	salmonellosis (cases)	all	surveillance (health system based)	1990-2005 (16)	Australia	Brisbane; Townsville	multi-site (city)	individual
82	Aminharati (2018)	shigellosis/bac. Dys. (incidence)	all	surveillance (health system based)	2012-2015 (4)	Iran	Yazd province	province	individual
83	Cheng (2017)	shigellosis/bac. Dys. (incidence)	all; <15; 15-64; ≥65	surveillance (health system based)	2006-2012 (7)	China	Hefei	city	individual
84	Gao (2014)	shigellosis/bac. Dys. (cases)	all	surveillance (health system based)	2004-2010 (7)	China	Changsha City	city	individual
85	Hao (2019)	shigellosis/bac. Dys. (cases)	<5; ≥5	surveillance (health system based)	2010-2015 (6)	China	Anhui province	province	individual
86	Lee (2017)	shigellosis/bac. Dys. (incidence)	all	national health datasets	1999-2013 (15)	Vietnam	Kon Tum Province	province	individual
87	Li (2013)	shigellosis/bac. Dys. (cases)	all	surveillance (health system based)	2006-2011 (6)	China	Wuhan	city	metro sub-districts
88	Li (2014)	shigellosis/bac. Dys. (incidence)	all	surveillance (health system based)	2006-2012 (7)	China	Guangzhou	city	individual
89	Li (2016)	shigellosis/bac. Dys. (incidence)	<6; 6-14; <15	surveillance (health system based)	2006-2012 (7)	China	Hefei, Anhui	city	individual
90	Li (2019b Weather)	shigellosis/bac. Dys. (cases)	all	surveillance (health system based)	2005-2011 (7)	China	Xiangxi	prefecture	individual
91	Liu (2019a STOTEN)	shigellosis/bac. Dys. (cases)	all	surveillance (health system based)	2005-2013 (9)	China	Jinan	city	individual
92	Liu (2020)	shigellosis/bac. Dys. (cases)	all; <6; ≥6	surveillance (health system based)	2014-2016 (3)	China	nationwide (316 cities)	country (city)	individual
93	Song (2018)	shigellosis/bac. Dys. (incidence)	all, 0-2, 3-6, 7-17, 18-64, ≥65	surveillance (health system based)	2002-2010 (9)	South Korea	nationwide (7 cities, 9 provinces)	country	province
94	Xu (2018)	shigellosis/bac. Dys. (incidence)	all	surveillance (health system based)	2010-2015 (6)	China	Hunan Province (122 counties)	province	county
95	Yan (2017)	shigellosis/bac. Dys. (incidence)	all	surveillance (health system based)	1970-2012 (forecast: 2005-2012) (43)	China	Beijing	city	individual

#	First author (Year)	A Etiology (outcome measure)[1]	B Ages (years)	C Outcome data source[2]	D Study period (years)	E Country[3]	F Location(s)	G Spatial coverage	H Model UOA[4]
96	Zhang (2007)	shigellosis/bac. Dys. (incidence)	all	hospital	1987-2000 (Jinan, 8); 1996-2003 (Baoan, 14)	China	Jinan; Baoan	multi-site (city)	individual
97	Zhang (2021)	shigellosis/bac. Dys. (incidence)	all	surveillance (health system based)	2013-2017 (5)	China	nationwide (N & S regions)	country (2 regions)	province
98	Dewan (2013)	typhoid fever (cases)	all	hospital	2005-2009 (5)	Bangladesh	Dhaka	city	individual
77	Thindwa (2019)	typhoid fever (incidence)	all	hospital	2011-2015 (5)	Malawi	Blantyre	city	individual

Notes:
[1]bac. Dys. = bacillary dysentery; [2]IDP = internally displaced persons; [3]LMICS = low and middle income countries; [4]UOA = unit of analysis

TABLE 5. (Part II: A, I-M) Review of studies of the effects of climate on diarrheal disease outcomes in the epidemiological literature, highlighting six of thirteen methodological issues discussed, arranged by diarrheal or gastroenteritis etiology (all-cause, campylobacteriosis, cholera, cryptosporidiosis and giardiasis, *E. coli* enteritis, norovirus, rotavirus, salmonellosis, shigellosis/bacillary dysentery (bac. dys.), typhoid fever).

#	First author (Year)	A Etiology (outcome measure) [1]	I Model time step (Study design)	J Statistical model(s) [2]	K Weather data source(s) [3]	L Meteorological variables [4]	M Covariates [5]
1	Alexander (2018)	all-cause (cases)	weekly (time series)	NBRM	station(s)	T, P	river level
2	Azage (2017)	all-cause (cases)	monthly (time series)	NBRM	GCD (8x8-km); reanalysis [LDAS] (1.25°)	T, P, RH	-
3	Bhavnani (2014)	all-cause (15-day prevalence)	monthly (each time for 15 days) (cohort; serial case-control)	logistic regression	rain gauge(s)	P	demographic (HH child <5, HH pop), SES (ownership), W&S
4	Boithias (2016)	all-cause (incidence)	weekly (time series)	NBRM	station(s); rain gauge(s); temp. probe	T, P	autocorrelation, stream discharge
5	Busch (2019)	all-cause (cases)	four times (repeated cross-section)	multilevel logistic GLMM	GCD (8x8-km)	P	SES (wealth), indigenous identity
6	Bush (2014a IJERPH)	all-cause (cases)	daily (time series)	Poisson GLM (2-stage)	station(s)	P, AT	trend, seasonality, DOW, beach closure
7	Bush (2014b EHP)	all-cause (cases)	daily (time series)	logistic regression; GAM with DLM	station(s)	P, AT	trend, season, DOW
8	Carlton (2014)	all-cause (incidence)	weekly (time series)	Poisson GLMM	rain gauge(s)	P	autocorrelation, hygiene, sanitation, social cohesion, remoteness
9	Checkley (2000)	all-cause (cases)	daily (time series)	Poisson GAM	station(s)	T, RH	autocorrelation, seasonality, El Niño
10	Chen (2018)	all-cause (incidence)	weekly (time series)	Poisson GLM with DLNM	station(s)	T, RH, WS, visibility	seasonality, trend, population
11	Chou (2010)	all-cause (incidence)	monthly (time series)	Poisson GLM	station(s)	T, P, RH	autocorrelation, trend, seasonality
12	Epstein (2020)	all-cause (cases)	3 panel waves; annual (repeated cross-section)	logistic regression (with restricted cubic splines)	reanalysis (0.05°); GCD (0.5°)	P	month, demographics (gender, age), breastfeeding status, urban residence, SES, W&S, remoteness, restricted cubic splines (nonlinearity)
13	Fang (2019)	all-cause (prevalence)	monthly (time series)	Poisson GAM	station(s)	T, P, RH, sunshine	seasonality, trend
14	Gao (2020)	all-cause (prevalence)	daily (time series)	Poisson GLM with DLNM	station(s)	T, P, RH	seasonality, trend, DOW
15	Gleason (2017)	all-cause (cases)	daily (time-stratified, bi-directional, case-crossover)	logistic regression (conditional); case-crossover (time-stratified, bi-directional)	station(s)	T, P, RH	strata: season, water source, demographics (gender, age, race), SES
16	Hashizume (2007)	all-cause (incidence)	weekly (time series)	Poisson GLM with DLM	station(s)	T, P	autocorrelation, seasonality, trend, holidays, river level
17	Horn (2018)	all-cause (cases)	weekly (time series)	Poisson GLM with DLM	reanalysis (0.05°); GCD (0.5°)	T, P	seasonality, trend, region/district

#	First author (Year)	A Etiology (outcome measure) [1]	I Model time step (Study design)	J Statistical model(s) [2]	K Weather data source(s) [3]	L Meteorological variables [4]	M Covariates [5]
18	Jagai (2015)	all-cause (incidence)	daily (time series)	Poisson GLM with DLM	station(s)	T, P	trend, strata: CSOs, demographics (age), season
19	Kulinkina (2016)	all-cause (cases)	weekly (cohort)	Poisson GLM	station(s)	T, P	seasonality, trend; strata: urban/rural
20	Lake (2009)	all-cause (incidence)	weekly (time series)	OLS (outcome: detrended residuals)	station(s)	T	autocorrelation, seasonality, trend, holidays, foreign travel cases by pathogen
21	Lama (2004)	all-cause (cases)	monthly (time series)	multiple linear regression	station(s)	T	El Niño, cholera
22	Lin (2016)	all-cause (cases)	daily (time series)	Poisson GAM	station(s)	T, P	trend, seasonality, DOW, holidays
23	Lloyd (2007)	all-cause (incidence)	monthly; study period (time series)	log-linear regression (meta-analysis)	GCD (0.5°)	T, P	SES (GDP or GCP), W&S, demographics (age), rural/urban
24	Mclver (2016)	all-cause (cases)	monthly (time series)	NBRM	station(s)	T, P	seasonality, trend
25	Mertens (2019)	all-cause (7-day prevalence)	weekly (cohort)	multi-level log-binomial GLMM; multi-level log-binomial GAMM	station(s); rain gauge	T, P	autocorrelation, village (RE), household, demographics (HH population, child sex, child age, maternal age), breastfeeding status, intervention group, SES (literacy, education, employment status, HH assets, HH ownership, caste), WaSH, group participation (e.g., community, credit finance, agriculture) [up to 1 covariate per 10 outcomes from expanded list]
26	Morral-Puigmal (2018)	all-cause (incidence)	daily (time series)	Poisson GLM with DLNM	station(s)	T, P	seasonality, trend, DOW, holidays, strata: demographics (sex, age), diagnosis group, climatic region, period
27	Musengimana (2016)	all-cause (incidence)	weekly (time series)	Poisson GLMM	station(s)	T	autocorrelation, sub-district
28	Onozuka (2010)	all-cause (cases)	weekly (time series)	Poisson GLM with DLM	station(s)	T, RH	autocorrelation, seasonality, trend
29	Phung (2015a IJB)	all-cause (cases)	weekly (time series)	Poisson GLM with DLM	station(s)	T, P, RH	seasonality, trend
30	Seidu (2013)	all-cause (cases)	biweekly (cohort)	Poisson GLM	station(s)	T, P	autocorrelation, seasonality
31	Singh (2001)	all-cause (cases)	annual (cross-section)	multiple linear regression	reanalysis (2.5°)	T, P	seasonality, SES, W&S
31	Singh (2001)	all-cause (cases)	monthly (time series)	Poisson GLM	reanalysis (2.5°)	T, P	seasonality, trend
32	Thompson (2015)	all-cause (incidence)	monthly (time series)	Poisson GLMM	station(s)	T, P, RH	district elevation, seasonality (month RE), district (RE), hospital distance, river level
33	H. Wang (2019)	all-cause (incidence)	daily (time series)	Poisson GLM with DLNM	station(s)	T, P, RH, DTR, AP, WS	seasonality, trend, DOW; strata: demographics (age)
34	P. Wang (2021)	all-cause (incidence)	daily (time series)	Poisson GLM with DLNM (meta-analysis)	weather stations	T, P, RH	autocorrelation, seasonality, trend, SES (GDP), demographic (population, education), health
35	Wangdi & Clements (2017)	all-cause (incidence)	monthly (time series)	RE Bayesian Poisson model	station(s)	T, P	demographics (age, sex)
36	Wu (2014)	all-cause (prevalence)	monthly (cohort)	logistic regression	station(s)	T, P	SES, demographics (number of children), flood-control status
37	Xu (2013)	all-cause (cases)	daily (time series)	Poisson GLM with DLNM	station(s)	T, RH, DTR	seasonality, trend, demographics
38	Xu (2014)	all-cause (cases)	daily (time series)	Poisson GLM with DLNM	reanalysis [LDAS] (6x6km)	T, P, RH	seasonality, trend, DOW, holidays
39	Zhou (2013)	all-cause (cases)	daily (time series)	Poisson GAM	station(s)	T, P, RH	seasonality, trend, DOW

#	First author (Year)	A Etiology (outcome measure) [1]	I Model time step (Study design)	J Statistical model(s) [2]	K Weather data source(s) [3]	L Meteorological variables [4]	M Covariates [5]
40	Allard (2011)	campylobacteriosis (incidence)	weekly (time series)	NBRM	station(s)	T	autocorrelation, seasonality, trend, holidays
41	Bi (2008)	campylobacteriosis (cases)	weekly (time series)	Poisson GLM	station(s)	T, P, RH	autocorrelation, season, trend
42	Fleury (2006)	campylobacteriosis (cases)	weekly (time series)	Poisson GLM; Poisson GAM	station(s)	T	autocorrelation, seasonality, trend, holidays, health region
20	Lake (2009)	campylobacteriosis (incidence)	weekly (time series)	OLS (outcome: detrended residuals)	station(s)	T	autocorrelation, seasonality, trend, holidays, foreign travel cases by pathogen
43	Milazzo (2017)	campylobacteriosis (incidence)	daily (time series)	Poisson GLM	station(s)	T	autocorrelation, trend, DOW, holidays
44	White (2009)	campylobacteriosis (incidence)	daily; weekly (time series; case-crossover)	Poisson GLM; case-crossover (time-stratified 2:1 matched)	station(s)	T, RH	seasonality, trend, river temperature
45	Ali (2013)	cholera (incidence)	monthly (time series)	SARIMA	station(s); gridded (4x4-km)	T, P, SST	seasonality, autocorrelation
46	de Magny (2008)	cholera (cases)	monthly (time series)	Poisson GLM with DLM	GCD (1°); GCD (2.5°)	P, SST	CHL-a, seasonality
47	Eisenberg (2013)	cholera (incidence)	daily; weekly (time series; case-crossover; dynamic modeling)	Poisson GLM with DLNM; case-crossover (conditional logistic regression); dynamic ("SIWR") model	stations; rain gauges; GCD (0.25°)	T, P, RH (t.s.); P (case-crossover)	study week
48	Hashizume (2008a Epi)	cholera (cases)	weekly (time series)	Poisson GLM with DLM	station(s)	T, P	autocorrelation, seasonality, trend, holidays, river level
49	Huq (2005)	cholera (incidence)	biweekly (time series)	Poisson GLM with DLM	unknown (presumably station(s))	T, P	water temp, probe, copepods, fecal coliforms, salinity, conductivity, water depth, dissolved O2 (varied by site)
50	Luque Fernández (2009)	cholera (cases)	weekly (time series)	Poisson GLM	station(s)	T, P	autocorrelation, seasonality, trend
51	Matsuda (2008)	cholera (cases)	monthly (time series)	autoregression (AR) model	station(s)	T, P	autocorrelation
52	Mendelsohn (2008)	cholera (incidence)	monthly (time series)	multiple linear regression	GCD (2.5°); GCD (9x9-km)	P, SST	SSH, CHL-a
53	Paz (2009)	cholera (cases)	annual (time series)	Poisson GLM	reanalysis (2.5°; GCD (5°)	T, SST, T anomoly, SST anomoly	autocorrelation
54	Reyburn (2011)	cholera (cases)	monthly (time series)	SARIMA	station(s)	T, P	autocorrelation, seasonality
55	Trærup (2011)	cholera (cases & fatalities)	annual (panel)	NBRM	station(s)	T, P	-
55	Trærup (2011)	cholera (cases)	monthly (time series)	NBRM	station(s)	T, P	drought, trend
55	Trærup (2011)	cholera (cases & fatalities)	annual (time series)	NBRM	station(s)	T, P	SES, W&S, population growth, demographic, cassava production
56	Wu (2018)	cholera (cases)	daily (case-crossover)	case-crossover (conditional logistic regression)	station(s)	T, P (n.s.), heatwave	heatwave; strata: P, demographics (sex, age), tree cover

#	First author (Year)	A Etiology (outcome measure) [1]	I Model time step (Study design)	J Statistical model(s) [2]	K Weather data source(s) [3]	L Meteorological variables [4]	M Covariates [5]
57	Chhetri (2017)	cryptosporidiosis and giardiasis (cases)	weekly (time series)	Poisson GLM with DLNM	GCD (10x10-km)	T (n.s.), P	trend, seasonality, holidays, pop. Growth (log(time)), dry/wet periods
58	Bifolchi (2014)	*E. coli* enteritis (incidence)	monthly (time series)	zero-inflated NBRM	station(s)	T	latitude, animal farming (horse, beef)
42	Fleury (2006)	*E. coli* enteritis (cases)	weekly (time series)	Poisson GLM; Poisson GAM	station(s)	T	autocorrelation, seasonality, trend, holidays, health region
59	Philipsborn (2016)	*E. coli* enteritis (incidence)	monthly (time series)	Poisson GLM ; Poisson GEE (meta-regression)	station(s); GCD (0.5°)	T, P	autocorrelation, mortality category, study
60	Lopman (2009)	norovirus (incidence)	daily; weekly (time series)	Poisson GLM (meta-analysis)	station(s)	T, RH, P (n.s.)	autocorrelation, trend (epidemic seasons), population immunity, improving diagnostics, seasonality, holidays, weekends, population density
61	P. Wang (2018b STOTEN)	norovirus (cases)	daily (time series)	NBRM with DLNM; NB GAM with DLNM	station(s)	T, P, RH, WS, solar rad.	Autocorrelation, seasonality, trend, DOW, holidays
62	Atchison (2010)	rotavirus (incidence)	weekly (time series)	Poisson GLM (meta-analysis); Poisson GLM with DLM (meta-analysis)	station(s)	T, P, RH	seasonality, trend, holidays
63	Celik (2015)	rotavirus (proportion; cases)	monthly (time series)	OLS; NBRM	station(s)	T, RH	autocorrelation, temperature cut-off indicator
64	D'Souza (2008)	rotavirus (cases)	weekly (time series)	log-linear regression; NBRM	station(s)	T, RH	seasonality, trend, autocorrelation, population
65	Hashizume (2008b E&I)	rotavirus (cases)	weekly (time series)	Poisson GLM	station(s)	T, RH	seasonality, trend, holidays, river level
66	Hervás (2014)	rotavirus (cases)	weekly (time series)	log-linear regression	station(s)	T, AP, VP, WS, solar rad.	-
67	Jagai, Sarkar, et al. (2012)	rotavirus (z-score)	monthly (time series)	GLMM (meta-analysis)	station(s)	T, P	NDVI, population, specific study
68	Levy (2009)	rotavirus (incidence)	monthly (time series)	RE GLS (random effects generalized least squares model); Poisson GEE (meta-analysis)	station(s)	T, P, RH	autocorrelation, trend, disease frequency, study
61	P. Wang (2018b STOTEN)	rotavirus (cases)	daily (time series)	NBRM with DLNM; NB GAM with DLNM	station(s)	T, P, RH, WS, solar rad.	Autocorrelation, seasonality, trend, DOW, holidays
69	Aik (2018)	salmonellosis (incidence)	weekly (time series)	NBRM with DLM	station(s)	T, P, RH	autocorrelation, trend, seasonality, holidays, period salmonellosis became legally notifiable
70	Britton (2010)	salmonellosis (cases)	monthly (time series)	NBRM with DLM	GCD (0.5°)	T	outbreak indicator, autocorrelation
71	D'Souza (2004)	salmonellosis (incidence)	monthly (time series)	Poisson GLM	station(s)	T, RH	seasonality, trend, outbreak, population
42	Fleury (2006)	salmonellosis (cases)	weekly (time series)	Poisson GLM; Poisson GAM	station(s)	T	autocorrelation, seasonality, trend, holidays, health region
72	Grjibovski (2013)	salmonellosis (cases)	monthly (time series)	multiple linear regression; NBRM	station(s)	T, P	autocorrelation, seasonality, trend, autocorrelation, population
73	Grjibovski (2014)	salmonellosis (cases)	monthly (time series)	NBRM; hockey-stick (n.s.)	station(s)	T, P	autocorrelation, seasonality, trend
74	Kovats (2004)	salmonellosis (cases)	weekly; biweekly; monthly by country (time series)	Poisson GLM with DLM	station(s)	T	autocorrelation, seasonality, trend, holidays

#	First author (Year)	A Etiology (outcome measure) [1]	I Model time step (Study design)	J Statistical model(s) [2]	K Weather data source(s) [3]	L Meteorological variables [4]	M Covariates [5]
20	Lake (2009)	salmonellosis (incidence)	weekly (time series)	OLS (outcome: detrended residuals)	station(s)	T	autocorrelation, seasonality, trend, holidays, foreign travel cases by pathogen
75	Lal (2016)	salmonellosis (incidence)	weekly (time series)	Poisson GLM with DLNM	station(s)	T	seasonality, trend
76	Milazzo (2016)	salmonellosis (incidence)	daily (time series)	Poisson GLM	station(s)	T	autocorrelation, trend, DOW, holidays
77	Thindwa (2019)	salmonellosis (incidence)	monthly (time series)	Poisson GLM with DLNM	station(s)	T, P	seasonality, trend
78	P. Wang (2018a EI)	salmonellosis (incidence)	daily (time series)	NB GAM with DLNM	station(s)	T, P, RH	seasonality, trend, holidays, DOW
79	Yun (2016)	salmonellosis (incidence)	weekly (time series)	NBRM	station(s)	T	trend
80	Zhang (2008a IJB)	salmonellosis (cases)	weekly (time series)	SARIMA	station(s)	T, P, RH	autocorrelation, seasonality, trend
81	Zhang (2010)	salmonellosis (cases)	weekly (Brisbane); monthly (Townsville) (time series)	Poisson GLM	station(s)	T, P	autocorrelation, seasonality, trend
82	Aminharati (2018)	shigellosis/bac. Dys. (incidence)	monthly (time series)	Poisson GLMM	station(s)	T, P, RH	dust condition, seasons, months
83	Cheng (2017)	shigellosis/bac. Dys. (incidence)	daily (time series)	Poisson GLM with DLNM; NBRM with DLNM	station(s)	T, RH	seasonality, trend, DOW
84	Gao (2014)	shigellosis/bac. Dys. (cases)	monthly (time series)	ARIMAX	station(s)	T, P (n.r.), RH, AP, WS (n.r.)	autocorrelation, trend
85	Hao (2019)	shigellosis/bac. Dys. (cases)	weekly (time series)	Poisson GLM with DLNM (meta-regression)	station(s)	T	autocorrelation, seasonality, trend
86	Lee (2017)	shigellosis/bac. Dys. (incidence)	monthly (time series)	RE NBRM	unknown	T, P, RH	month
87	Li (2013)	shigellosis/bac. Dys. (cases)	daily (time series)	NBRM; Poisson GAM	station(s)	T, P, RH, WS	seasonality, trend, DOW, holidays
88	Li (2014)	shigellosis/bac. Dys. (incidence)	monthly (time series)	NBRM	station(s)	T, P, RH, AP, WS, sunshine	trend
89	Li (2016)	shigellosis/bac. Dys. (incidence)	daily (time series)	Poisson GLM with DLNM	station(s)	T, RH	seasonality, trend, DOW
90	Li (2019b Weather)	shigellosis/bac. Dys. (cases)	monthly (time series)	Poisson GLM	station(s)	T, P, RH, AP, WS, sunshine	autocorrelation, seasonality, trend
91	Liu (2019a STOTEN)	shigellosis/bac. Dys. (cases)	daily (time series)	Poisson GAM; Poisson GAM with DLNM	station(s)	T, P, RH, WS, sunshine	seasonality, trend
92	Liu (2020)	shigellosis/bac. Dys. (cases)	daily (time series)	Poisson GLM with DLNM	station(s)	T, P, RH	autocorrelation, seasonality, trend
93	Song (2018)	shigellosis/bac. Dys. (incidence)	weekly (time series)	Poisson GAM	GCD (1x1-km)	T, P	trend, seasonality; strata: demographics (age, gender), season
94	Xu (2018)	shigellosis/bac. Dys. (incidence)	monthly (time series)	BSTHM	station(s)	T, P, RH, AP, WS, sunshine	trend, county, ratio of county risk to overall risk
95	Yan (2017)	shigellosis/bac. Dys. (incidence)	monthly (time series)	ARIMAX	station(s)	T, P, VP, WS, RH (n.s.)	seasonality, autocorrelation

#	First author (Year)	A Etiology (outcome measure) [1]	I Model time step (Study design)	J Statistical model(s) [2]	K Weather data source(s) [3]	L Meteorological variables [4]	M Covariates [5]
96	Zhang (2007)	shigellosis/bac. Dys. (incidence)	monthly (time series)	SARIMA	station(s)	T, P, RH, AP	autocorrelation, seasonality, trend
97	Zhang (2021)	shigellosis/bac. Dys. (incidence)	monthly (time series)	BSTHM; GeoDetector	station(s)	T, P, RH, AP, sunshine	population density, WW discharge, health beds, health technicians, SES (GDP, illiteracy)
98	Dewan (2013)	typhoid fever (cases)	weekly (time series)	Poisson GLM	station(s)	T, P	seasonality, trend, holidays, river level
77	Thindwa (2019)	typhoid fever (incidence)	monthly (time series)	Poisson GLM with DLNM	station(s)	T, P	seasonality, trend

Notes:

[1] bac. Dys. = bacillary dysentery

[2] OLS = ordinary least squares; NB = negative binomial; NBRM = negative binomial regression model; GL(M)M = generalized linear (mixed) model; GA(M)M = generalized additive (mixed) model; GLS = generalized least squares; GEE = generalized estimating equation; AR = autoregressive or autoression; RE = random effects; FE = fixed effects; (S)ARIMA(X) = (seasonal) autoregressive integrated moving average (with exogenous variables); BSTHM = Bayesian space-time hierarchical model

[3] station(s) = weather station(s); GCD = gridded climate data product; reanalysis = climated reanalysis product; (G)LDAS = (global) land data assimilation system.

[4] T = ambient temperature; DTR = diurnal temperature range; P = precipitation or rainfall; RH = relative humidity; AT = apparent temperature; VP = (water) vapor pressure; AP = air or atmospheric pressure; solar rad. = solar radiation; sunshine = sunshine duration (e.g., hours); SST = sea surface temperature; SSH = sea-surface height; CHL-a = chlorophyll-a; [4] n.r. indicates not reported; n.s. indicates not significant

[5] DOW = day-of-week; NDVI = Normalized Difference Vegetation Index; W&S = water and sanitation' WASH = water, sanitation, and hygiene; WW = wastewater; CHL-a = chlorophyll-a

TABLE 6. Model summary with references, R packages, and example studies for generalized linear models (GLM), generalized additive models (GAM), generalized estimating equations (GEE), (seasonal) autoregressive integrated moving average (with exogenous variables) ((S)ARIMA(X)), distributed lag (non-linear) models (DL(N)M), case-crossover analysis, and four data mining techniques (multivariate adaptive regression splines (MARS), classification and regression trees (CART), Bagging CART (BC), Random Forest (RF)).

Model (References)	R packages	Diarrheal Disease & Weather Studies
Time Series Models		
GLM (Cameron & Trivedi, 1998, 2013; McCullagh & Nelder, 1989)	glm: stats::glm() or lme4::glmer() with random effects negative binomial glm: MASS::glm.nb() or lme4::glmer.nb() with random effects hurdle model: pscl::hurdle() zero-inflated model: pscl::zeroinf()	*GLM:* (Bi et al., 2008; Chou et al., 2010; de Magny et al., 2008; El-Fadel et al., 2012; Fleury et al., 2006; Galway et al., 2015; Hashizume, Armstrong, Hajat, et al., 2008; Hashizume, Armstrong, Wagatsuma, et al., 2008; Hashizume et al., 2007; Hashizume, Wagatsuma, et al., 2008; He et al., 2006; Hsieh et al., 2015; Hu et al., 2007; Kovats, Edwards, Hajat, Armstrong, Ebi, Menne, Cowden, et al., 2004; Lippmann et al., 2013; Naumova et al., 2000, 2005, 2007; Onozuka et al., 2010; Shortridge & Guikema, 2014) *quasi-Poisson :* (de Magny et al., 2008; Xu et al., 2014; Zhou et al., 2013) *GLM negative binomial (NB) :* (Brubacher et al., 2020; D'Souza et al., 2004; Grjibovski et al., 2013; Hashizume et al., 2011; D. Lee et al., 2019; Onozuka & Hashizume, 2011; Tam, Rodriguez, et al., 2006) *Autoregressive-adjusted Poisson:* (Bi et al., 2008; Hashizume, Armstrong, Hajat, et al., 2008; Hashizume et al., 2007; Onozuka et al., 2010; Y. Zhang et al., 2008a, 2010) *Poisson harmonic regression :* (Butler, 2013; Checkley et al., 2000; Hashizume, Armstrong, Hajat, et al., 2008; Jagai, Griffiths, et al., 2012; Nygård et al., 2004)
GAM (Hastie & Tibshirani, 1986, 1990a; Wood, 2004, 2006, 2017)	gam (Chambers & Hastie, 1991; Hastie, 2020; Hastie & Tibshirani, 1990c) mgcv (Wood, 2001, 2020)	*GAM:* (Baccini et al., 2007; Beaudeau, Le Tertre, et al., 2012; Beaudeau, Zeghnoun, et al., 2014; Beaudeau, Schwartz, et al., 2014; Bush et al., 2014b; Fleury et al., 2006; Galway et al., 2015; Shortridge & Guikema, 2014; Tornevi et al., 2013, 2014, 2015; Uejio et al., 2014; Zhou et al., 2013) *GAM negative binomial (NB):* (Martinez-Urtaza et al., 2008)
GEE (K.-Y. Liang & Zeger, 1986; Zeger & Liang, 1986)	gee (Carey, 2019) geepack (Halekoh et al., 2006)	*GEE : 59, 68* (K. Levy, Hubbard, Eisenberg, 2009; Philipsborn et al., 2016)

Model (References)	R packages	Diarrheal Disease & Weather Studies
ARIMA/ARIMAX/ SARIMA/SARIMAX (Box & Jenkins, 1976; Brockwell & Davis, 2002; Chatfield, 2001; Cryer & Chan, 2008; Shumway & Stoffer, 2006, 2017; Tobías & Saez, 2004)	stats::arima() sarima::sarim a()	*ARMA* : (Drayna et al., 2010) *ARIMA* : (Redman et al., 2007) *ARIMA*X: (Chadsuthi et al., 2012a ; Gao et al., 2014; Wangdi et al., 2010) *SARIMA*: (Hu et al., 2007; Wangdi et al., 2010b; Y. Zhang et al., 2007, 2008a, 2008b)
DLNM/ DLM DLM (Almon, 1965; Schwartz, 2000b; Zanobetti et al., 2000) DLNM (Gasparrini, 2011; Gasparrini et al., 2010, 2017; Gasparrini & Leone, 2014)	dlnm::dlnm()	*DLM*: (Hall et al., 2011; Hashizume, Armstrong, Hajat, et al., 2008b; Hsieh et al., 2015; Jagai et al., 2015; Onozuka & Hashizume, 2011) *DLN*M: (Chhetri et al., 2017; M. C. Eisenberg et al., 2013; Tornevi et al., 2013, 2015; Y. Wang et al., 2019; Xu et al., 2014)
Case-crossover Analysis (Maclure, 1991; Maclure & Mittleman, 2000; Mittleman et al., 1995; Mittleman & Mostofsky, 2014; Perrakis et al., 2014) See modifications in Perrakis *et al.* (2014 refs. 11-21)	N/A	case-crossover: (Ding et al., 2013; M. C. Eisenberg et al., 2013 ; Gleason & Fagliano, 2017; Jagai et al., 2017; C. J. Lin et al., 2015; Nichols et al., 2009; K. M. Thomas et al., 2006)
Data Mining Techniques		
MARS (Friedman, 1991)	earth (Milborrow, 2011, 2020) *Alternative*: Salford System (Steinberg et al., 1999)	(Shortridge & Guikema, 2014)
CART (a.k.a. RTA) (Breiman et al., 1984; Clark & Pregibon, 1992; Verbyla, 1987)	rpart (Therneau & Atkinson, 1997, 2019)	(Guan et al., 2008; Hu, Mengersen, & Tong, 2010; Hu, Mengersen, Fu, et al., 2010; Shortridge & Guikema, 2014; Xu et al., 2015)
BC (a.k.a. BT) (Hastie et al., 2001, 2009)		(Shortridge & Guikema, 2014)
RF Random Forest (Liaw & Wiener, 2002)	randomForest (Liaw & Wiener, 2002)	(Shortridge & Guikema, 2014)

GLM, generalized linear model; **GAM**; generalized additive model; **ARIMA**, autoregressive integrated moving average; **ARIMAX**, autoregressive Integrated Moving Average with Exogenous Variables; **ARMA**, Autoregressive Moving Average; **SARIMA**, Seasonal Autoregressive Integrated Moving Average; DLM, distributed lag model; non-linear distributed lag model; **MDLNM**, multilevel distributed lag model; **MARS**, Multivariate adaptive regression splines; **CART**, Classification and Regression Tree; **RTA**, Regression Tree Analysis; **BC**, Bagged CART; **BT**, Bagging Trees; **RF**, Random Forest

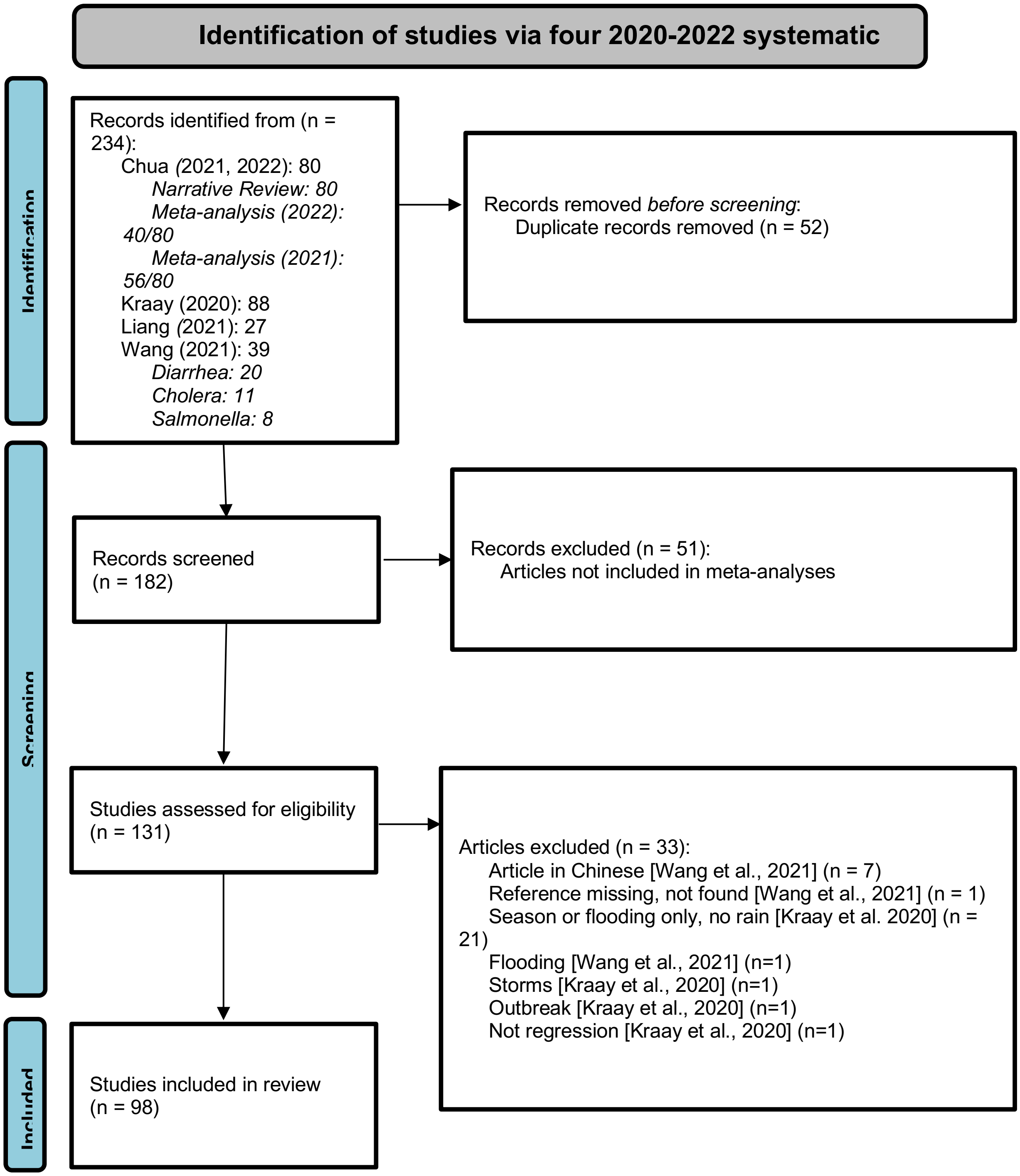
TABLE 7. Strengths and weaknesses of different types of weather data products: weather station (S) data, gridded weather or climate data (GWD or GCD) products, and data assimilation (DA) products (a.k.a. reanalyses). ***Source:*** **adapted from Auffhammer** ***et al.*** (2013c)**.**

Data	Strengths	Weaknesses	Example datasets
Station (S)	**Local high-resolution data in many locations:** • Global network of weather stations (since ~1850) and satellites (first launched in 1960) to measure and record weather outcomes • Free daily data available • Numerous variables available: T (min, max), PPT (total), snowfall, snow depth, etc.	**Incomplete records:** • Weather data is proprietary and expensive in some countries, limiting availability • Birth and death of weather stations • Missing observations over time • 2/3 of GHCN weather stations only report precipitation **Weather stations may not exist in all locations of interest:** • High variation in spatial and temporal weather station coverage across the globe • Higher income countries (e.g., U.S. & EU-15) have longer time series records and higher spatial density of weather stations • May need to do own spatial interpolation depending on the proximity/coverage of the study area with weather stations	**Global:** • NOAA NCDC & NCEI Ex: Global Historic Climatology Network (GHCN-Daily & GHCN-Monthly (https://www.ncdc.noaa.gov/cdo-web/search) **Regional:** • China: CDMSS China Meteorological Data Sharing Service System (http://data.cma.cn/)
Gridded (GCD, GWD)	**Complete spatial coverage of weather over land:** • Free data available • Easy to import into formats commonly used by researchers	**Limited spatio-temporal resolution:** • Spatial and temporal resolution (monthly, 0.5-degree grids) may be inadequate for some data needs at the global level (e.g., CRU & UDEL) or in places with sparse data **Potential biases due to interpolations:** • Interpolations over missing observations or areas where there are no weather stations in some grids	**Global:** • CRU (http://www.cru.uea.ac.uk/data/)[17, 23, 53, 59, 70] • UDEL (Mitchell & Jones, 2005) • [PPT] CMORPH (Joyce et al., 2004 ; Xie et al., 2019)[2] • [PPT] NASA TRMM TMPA (1997-April 2015); *current*: Integrated Multi-satellitE Retrievals for GPM (IMERG) (Huffman et al., 2019)[47, 56] • [PPT] PERSIANN-CDR (https://www.ncdc.noaa.gov/cdr/atmospheric/precipitation-persiann-cdr) **Regional:** • CONUS: Daymet (Thornton et al., 1997) • CONUS: PRISM (Daly et al., 1994, 2008; PRISM Climate Group, 2004) (http://www.prism.oregonstate.edu/)

Data	Strengths	Weaknesses	Example datasets
Data Assimilation (DA) (Reanalyses)[1]	**Complete spatial coverage at finer temporal resolution:** • Can get finer temporal/spatial resolution for some areas than are offered at a global scale (especially data-sparse locations) Combines use of physical models with observational data	**Output cannot be forced to perfectly match observational data:** • Limited resolution and influenced by general circulation model (GCM) even with observations • Models are imperfect and have systematic biases that may not always be corrected by observational data constraints **Comparative (dis)advantages depend on underlying data richness (sparsity):** • Likely worse than other products in data regions (i.e., U.S. & Europe) May be better option if data is sparse, but it still has limitations as a model prediction	**Global:** • NOAA NCEP-NCAR (R1) (Kalnay et al., 1996; Kistler et al., 2001)[31, 53] and NCEP Reanalysis (R2) • ECMWF ERA40, ERA-Interim (European Centre for Medium-Range Weather Forecasts (ECMWF), 2010) • NCC (NCEP/NCAR Corrected by CRU) (Ngo-Duc et al., 2005a, 2005b) • [PPT] CHIRPS (Funk 2015) (https://www.chc.ucsb.edu/data/chirps)[12, 17] **Global - Land Data Assimilation Systems (LDAS):** • MODIS Land surface temperature (LST) (https://modis.gsfc.nasa.gov/data/dataprod/)[38] • GLDAS (Rodell et al., 2004) (https://ldas.gsfc.nasa.gov/gldas)[2]

Note: [1] Refer to https://climatedataguide.ucar.edu/climate-data/reanalysis for descriptions, key strengths, key limitations, and data access information for reanalysis datasets and https://climatedataguide.ucar.edu/climate-data/atmospheric-reanalysis-overview-comparison-tables for a list of additional reanalyses datasets, including examples of which are 1st (NCEP-NCAR; NCEP-DOE), 2nd (ERA40; JRA25), and 3rd (ERA-Interim; MERRA; CFSR) generation.

2.7. Figures


Identification of studies via four 2020-2022 systematic
Identification
Screening
Included
Records identified from (n = 234):
Chua (2021, 2022): 80
Narrative Review: 80
Meta-analysis (2022): 40/80
Meta-analysis (2021): 56/80
Kraay (2020): 88
Liang (2021): 27
Wang (2021): 39
Diarrhea: 20
Cholera: 11
Salmonella: 8
Records removed before screening:
Duplicate records removed (n = 52)
Records screened
(n = 182)
Records excluded (n = 51):
Articles not included in meta-analyses
Studies assessed for eligibility
(n = 131)
Articles excluded (n = 33):
Article in Chinese [Wang et al., 2021] (n = 7)
Reference missing, not found [Wang et al., 2021] (n = 1)
Season or flooding only, no rain [Kraay et al. 2020] (n = 21)
Flooding [Wang et al., 2021] (n=1)
Storms [Kraay et al., 2020] (n=1)
Outbreak [Kraay et al., 2020] (n=1)
Not regression [Kraay et al., 2020] (n=1)
Studies included in review
(n = 98)

2.8. Supplementary tables and equations

TABLE 8. Summary table of study outcome(s) by diarrheal etiology, categorized by type of pathogen (bacteria, protozoa, viruses). Four of the 98 studies reported multiple outcomes.[20, 42, 61, 77]

Etiology (n = studies)	Studies (first author year)	#
all-cause* (39)	Alexander (2018); Azage (2017); Bhavnani (2014); Boithias (2016); Busch (2019); Bush (2014a IJERPH); Bush (2014b EHP); Carlton (2014); Checkley (2000); Chen (2018); Chou (2010); Epstein (2020); Fang (2019); Gao (2020); Gleason (2017); Hashizume (2007); Horn (2018); Jagai (2015); Kulinkina (2016); Lake (2009); Lama (2004); Lin (2016); Lloyd (2007); Mclver (2016); Mertens (2019); Morral-Puigmal (2018); Musengimana (2016); Onozuka (2010); Phung (2015a IJB); Seidu (2013); Singh (2001); Thompson (2015); H. Wang (2019); Wang (2021); Wangdi & Clements (2017); Wu (2014); Xu (2013); Xu (2014); Zhou (2013)	1-39
Bacteria (54)		
campylobacteriosis (6)	Allard (2011); Bi (2008); Fleury (2006); Lake (2009); Milazzo (2017); White (2009)	20, 40-44
cholera (12)	Ali (2013); de Magny (2008); Eisenberg (2013); Hashizume (2008a Epi); Huq (2005); Luque Fernández (2009); Matsuda (2008); Mendelsohn (2008); Paz (2009); Reyburn (2011); Trærup (2011); Wu (2018)	45-56
***E. coli* enteritis (3)**	Bifolchi (2014); Fleury (2006); Philipsborn (2016)	42, 58, 59
salmonellosis (15)	Aik (2018); Britton (2010); D'Souza (2004); Fleury (2006); Grjibovski (2013); Grjibovski (2014); Kovats (2004); Lake (2009); Lal (2016); Milazzo (2016); Thindwa (2019); Wang (2018a EI); Yun (2016); Zhang (2008a IJB); Zhang (2010)	20, 42, 69-81
shigellosis/ bacillary dysentery (16)	Aminharati (2018); Cheng (2017); Gao (2014); Hao (2019); Lee (2017); Li (2013); Li (2014); Li (2016); Li (2019b Weather); Liu (2019a STOTEN); Liu (2020); Song (2018); Xu (2018); Yan (2017); Zhang (2007); Zhang (2021)	82-97
typhoid fever (2)	Dewan (2013); Thindwa (2019)	77, 98
Protozoa (1)		
cryptosporidiosis and giardiasis (1)	Chhetri (2017)	57
Viruses (11)		
norovirus (2)	Lopman (2009); Wang (2018b STOTEN)	60, 61
rotavirus (8)	Atchison (2010); Celik (2015); D'Souza (2008); Hashizume (2008b E&I); Hervás (2014); Jagai, Sarkar, et al. (2012); Levy (2009); Wang (2018b STOTEN)	61- 68

**all-cause* includes all non-specified or general categories of diarrhea, gastroenteritis, and waterborne diseases, which vary in name by author, included but not limited to all-cause: diarrhea, infectious diarrhea, non-cholera diarrhea, acute gastrointestinal illness (AGI), gastroenteritis (GI), enteric illness, waterborne diseases, etc.

TABLE 9. Summary table of studies by continent and region

Continent (n = studies)	Studies (first author year)	#
Global (3)	Levy (2009); Lloyd (2007); Philipsborn (2016)	23, 59, 68
Asia (50)	Aik (2018); Alexander (2018); Ali (2013); Aminharati (2018); Boithias (2016); Bush (2014b EHP); Celik (2015); Chen (2018); Cheng (2017); Chou (2010); de Magny (2008); Dewan (2013); Fang (2019); Gao (2014); Gao (2020); Grjibovski (2013); Grjibovski (2014); Hao (2019); Hashizume (2007); Hashizume (2008a Epi); Hashizume (2008b E&I); Huq (2005); Kulinkina (2016); Lee (2017); Li (2013); Li (2014); Li (2016); Li (2019b Weather); Liu (2019a STOTEN); Liu (2020); Luque Fernández (2009); Matsuda (2008); Mclver (2016); Mertens (2019); Onozuka (2010); Phung (2015a IJB); Song (2018); Thompson (2015); Wang (2018a EI); Wang (2018b STOTEN); H. Wang (2019); P. Wang (2021); Wangdi & Clements (2017); Wu (2014); Wu (2018); Xu (2018); Yan (2017); Zhang (2007); Zhang (2021); Zhou (2013)	1, 4, 7, 10, 11, 13, 14, 16, 19, 24, 25, 28, 29, 32-36, 39, 45, 46, 48- 51, 56, 61, 63, 65, 69, 72, 73, 78, 82- 98
Africa (12)	Azage (2017); Busch (2019); Epstein (2020); Horn (2018); Jagai, Sarkar, et al. (2012); Mendelsohn (2008); Musengimana (2016); Paz (2009); Reyburn (2011); Seidu (2013); Thindwa (2019); Trærup (2011)	2, 5, 12, 17, 27, 30, 52, 53, 54, 55, 67, 77
Australia (9)	Bi (2008); D'Souza (2004); D'Souza (2008); Milazzo (2016); Milazzo (2017); Xu (2013); Xu (2014); Zhang (2008a IJB); Zhang (2010)	37, 38, 41, 43, 64, 71, 76, 80, 81
Europe (7)	Atchison (2010); Hervás (2014); Kovats (2004); Lake (2009); Lopman (2009); Morral-Puigmal (2018); Yun (2016)	20, 26, 60, 62, 66, 74, 79
North America (10)	Allard (2011); Bifolchi (2014); Bush (2014a IJERPH); Chhetri (2017); Eisenberg (2013); Fleury (2006); Gleason (2017); Jagai (2015); Lin (2016); White (2009)	6, 15, 18, 22, 40, 42, 44, 47, 57, 58
Oceania (3)	Britton (2010); Lal (2016); Singh (2001)	31, 70, 75
South America (4)	Bhavnani (2014); Carlton (2014); Checkley (2000); Lama (2004)	3, 8, 9, 21

TABLE 10. Summary table of studies by age group

Age group (n = studies)[1]	Studies (first author year)	#
≤2 (1)	Epstein (2020)	12
<5 (10)	Alexander (2018); Atchison (2010); Azage (2017); Celik (2015); D'Souza (2008); Hervás (2014); Musengimana (2016); Wang (2018b STOTEN); Wu (2014); Xu (2013)	1, 2, 27, 36, 37, 61-64, 66
<6 (2)	Mertens (2019)	25
<10 (2)	Thompson (2015)	32
<16 (1)	Busch (2019)	5
>13 (1)	Bush (2014a IJERPH)	6
≥18 (1)	Epstein (2020)	12
≥65 (1)	Alexander (2018); Atchison (2010); Azage (2017); Celik (2015); D'Souza (2008); Hervás (2014); Musengimana (2016); Wang (2018b STOTEN); Wu (2014); Xu (2013)	1, 2, 27, 36, 37, 61-64, 66
all ages (62)	Aik (2018); Ali (2013); Allard (2011); Aminharati (2018); Bhavnani (2014); Bi (2008); Bifolchi (2014); Boithias (2016); Britton (2010); Bush (2014b EHP); Carlton (2014); Chen (2018); Chhetri (2017); D'Souza (2004); de Magny (2008); Dewan (2013); Eisenberg (2013); Fleury (2006); Gao (2014); Gao (2020); Grjibovski (2013); Grjibovski (2014); Hashizume (2008a Epi); Hashizume (2008b E&I); Horn (2018); Huq (2005); Jagai, Sarkar, et al. (2012); Kovats (2004); Kulinkina (2016); Lake (2009); Lal (2016); Lee (2017); Li (2013); Li (2014); Li (2019b Weather); Lin (2016); Liu (2019a STOTEN); Lopman (2009); Luque Fernández (2009); Mclver (2016); Mendelsohn (2008); Milazzo (2016); Milazzo (2017); Onozuka (2010); Paz (2009); Philipsborn (2016); Phung (2015a IJB); Reyburn (2011); Seidu (2013); Thindwa (2019); Trærup (2011); Wang (2018a EI); Wang (2021); White (2009); Xu (2018); Yan (2017); Yun (2016); Zhang (2007); Zhang (2008a IJB); Zhang (2010); Zhang (2021); Zhou (2013)	3, 4, 7, 8, 10, 14, 17, 19, 20, 22, 24, 28, 29, 30, 34, 39, 40-50, 52-55, 57-60, 65, 67, 69-82, 84, 86, 87, 88, 90, 91, 94-98
multiple age groups (18)	Cheng (2017); Chou (2010); Fang (2019); Gleason (2017); Hao (2019); Hashizume (2007); Jagai (2015); Levy (2009); Li (2016); Liu (2020); Lloyd (2007); Morral-Puigmal (2018); Singh (2001); Song (2018); H. Wang (2019); Wangdi & Clements (2017); Wu (2018); Xu (2014)	11, 13, 15, 16, 18, 23, 26, 31, 33, 35, 38, 56, 68, 83, 85, 89, 92, 93

Notes: [1]Counts and references listed by age group are for studies with estimates for single age groups, excluding those that include multiple age groups.

TABLE 11. Summary of statistical model types and characteristics by study (in alphabetical order).

#	Study (First Author, Year)	Model(s)	multiple model types (n=18)	OLS (n=2)	multiple linear (n=3)	logistic (n=8)	binomial (n=1)	log-linear (n=3)	GAM (n=12)	Poisson GLM (n=47)	Poisson GAM (n=8)	Poisson GEE (n=2)	NBRM (n=19)	NB GAM (n=2)	GLMM (n=7)	GAMM (n=1)	random effects (RE) (n=4)	GLS (n=1)	Bayesian Poisson (n=1)	BSTHM (n=2)	auto-regressive (AR) (n=1)	(S)ARIMA(X) (n=6)	DLM (n=13)	DLNM (n=18)	zero-inflated (n=1)	case-crossover (n=4)	meta-regression/-analysis (n=8)
69	Aik (2018)	NBRM with DLM	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	1	-	-	-	-
1	Alexander (2018)	NBRM	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
45	Ali (2013)	SARIMA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-
40	Allard (2011)	NBRM	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
82	Aminharati (2018)	Poisson GLMM	-	-	-	-	-	-	-	1	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-
62	Atchison (2010)	Poisson GLM (meta-analysis); Poisson GLM with DLM (meta-analysis)	1	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	1
2	Azage (2017)	NBRM	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3	Bhavnani (2014)	logistic regression	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
41	Bi (2008)	Poisson GLM	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
58	Bifolchi (2014)	zero-inflated NBRM	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	1	-	-
4	Boithias (2016)	NBRM	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
70	Britton (2010)	NBRM with DLM	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	1	-	-	-	-
5	Busch (2019)	logistic GLMM	-	-	-	1	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-
6	Bush (2014a IJERPH)	Poisson GLM (2-stage)	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7	Bush (2014b EHP)	logistic regression; GAM with DLM	1	-	-	1	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-
8	Carlton (2014)	RE Poisson GLMM	-	-	-	-	-	-	-	1	-	-	-	-	1	-	1	-	-	-	-	-	-	-	-	-	-
63	Celik (2015)	OLS; NBRM	1	1	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
9	Checkley (2000)	Poisson GAM	-	-	-	-	-	-	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	Chen (2018)	Poisson GLM with DLNM	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-

#	Study (First Author, Year)	Model(s)	multiple model types (n=18)	OLS (n=2)	multiple linear (n=3)	logistic (n=8)	binomial (n=1)	log-linear (n=3)	GAM (n=12)	Poisson GLM (n=47)	Poisson GAM (n=8)	Poisson GEE (n=2)	NBRM (n=19)	NB GAM (n=2)	GLMM (n=7)	GAMM (n=1)	random effects (RE) (n=4)	GLS (n=1)	Bayesian Poisson (n=1)	BSTHM (n=2)	auto-regressive (AR) (n=1)	(S)ARIMA(X) (n=6)	DLM (n=13)	DLNM (n=18)	zero-inflated (n=1)	case-crossover (n=4)	meta-regression/-analysis (n=8)
83	Cheng (2017)	Poisson GLM with DLNM; NBRM with DLNM	1	-	-	-	-	-	-	1	-	-	1	-	-	-	-	-	-	-	-	-	-	1	-	-	-
57	Chhetri (2017)	Poisson GLM with DLNM	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-
11	Chou (2010)	Poisson GLM	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
71	D'Souza (2004)	Poisson GLM	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
64	D'Souza (2008)	log-linear regression; NBRM	1	-	-	-	-	1	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
46	de Magny (2008)	Poisson GLM with DLM	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-
98	Dewan (2013)	Poisson GLM	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
47	Eisenberg (2013)	Poisson GLM with DLNM; case-crossover (conditional logistic regression); dynamic ("SIWR") model	1	-	-	1	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	1	-
12	Epstein (2020)	logistic regression (with restricted cubic splines)	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
13	Fang (2019)	Poisson GAM	-	-	-	-	-	-	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
42	Fleury (2006)	Poisson GLM; Poisson GAM	1	-	-	-	-	-	1	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
84	Gao (2014)	ARIMAX	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-
14	Gao (2020)	Poisson GLM with DLNM	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-
15	Gleason (2017)	logistic regression (conditional); case-crossover (time-stratified, bi-directional)	1	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-
72	Grjibovski (2013)	multiple linear regression; NBRM	1	-	1	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
73	Grjibovski (2014)	NBRM; hockey-stick (n.s.)	1	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
85	Hao (2019)	Poisson GLM with DLNM (meta-regression)	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	1
16	Hashizume (2007)	Poisson GLM with DLM	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-

#	Study (First Author, Year)	Model(s)	multiple model types (n=18)	OLS (n=2)	multiple linear (n=3)	logistic (n=8)	binomial (n=1)	log-linear (n=3)	GAM (n=12)	Poisson GLM (n=47)	Poisson GAM (n=8)	Poisson GEE (n=2)	NBRM (n=19)	NB GAM (n=2)	GLMM (n=7)	GAMM (n=1)	random effects (RE) (n=4)	GLS (n=1)	Bayesian Poisson (n=1)	BSTHM (n=2)	auto-regressive (AR) (n=1)	(S)ARIMA(X) (n=6)	DLM (n=13)	DLNM (n=18)	zero-inflated (n=1)	case-crossover (n=4)	meta-regression/-analysis (n=8)
48	Hashizume (2008a Epi)	Poisson GLM with DLM	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-
65	Hashizume (2008b E&I)	Poisson GLM	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
66	Hervás (2014)	log-linear regression	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17	Horn (2018)	Poisson GLM with DLM	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-
49	Huq (2005)	Poisson GLM with DLM	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-
67	Jagai, Sarkar, et al. (2012)	GLMM (meta-analysis)	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	1
18	Jagai (2015)	Poisson GLM with DLM	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-
74	Kovats (2004)	Poisson GLM with DLM	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-
19	Kulinkina (2016)	Poisson GLM	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
20	Lake (2009)	OLS (outcome: detrended residuals)	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
75	Lal (2016)	Poisson GLM with DLNM	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-
21	Lama (2004)	multiple linear regression	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
86	Lee (2017)	RE NBRM	-	-	-	-	-	-	-	-	-	-	1	-	-	-	1	-	-	-	-	-	-	-	-	-	-
68	Levy (2009)	RE GLS (random effects generalized least squares model); Poisson GEE	1	-	-	-	-	-	-	-	-	1	-	-	-	-	1	1	-	-	-	-	-	-	-	-	1
87	Li (2013)	NBRM; Poisson GAM	1	-	-	-	-	-	1	-	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
88	Li (2014)	NBRM	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
89	Li (2016)	Poisson GLM with DLNM	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-
90	Li (2019b Weather)	Poisson GLM	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
22	Lin (2016)	Poisson GAM	-	-	-	-	-	-	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

#	Study (First Author, Year)	Model(s)	multiple model types (n=18)	OLS (n=2)	multiple linear (n=3)	logistic (n=8)	binomial (n=1)	log-linear (n=3)	GAM (n=12)	Poisson GLM (n=47)	Poisson GAM (n=8)	Poisson GEE (n=2)	NBRM (n=19)	NB GAM (n=2)	GLMM (n=7)	GAMM (n=1)	random effects (RE) (n=4)	GLS (n=1)	Bayesian Poisson (n=1)	BSTHM (n=2)	auto-regressive (AR) (n=1)	(S)ARIMA(X) (n=6)	DLM (n=13)	DLNM (n=18)	zero-inflated (n=1)	case-crossover (n=4)	meta-regression/-analysis (n=8)
91	Liu (2019a STOTEN)	Poisson GAM; Poisson GAM with DLNM	1	-	-	-	-	-	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-
92	Liu (2020)	Poisson GLM with DLNM	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-
23	Lloyd (2007)	log-linear regression	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
60	Lopman (2009)	Poisson GLM (meta-analysis)	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
50	Luque Fernández (2009)	Poisson GLM	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
51	Matsuda (2008)	autoregression (AR) model	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
24	McIver (2016)	NBRM	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
52	Mendelsohn (2008)	multiple linear regression	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
25	Mertens (2019)	log-binomial GLMM; log-binomial GAMM	1	-	-	-	1	-	1	-	-	-	-	-	1	1	-	-	-	-	-	-	-	-	-	-	-
76	Milazzo (2016)	Poisson GLM	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
43	Milazzo (2017)	Poisson GLM	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
26	Morral-Puigmal (2018)	Poisson GLM with DLNM	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-
27	Musengimana (2016)	Poisson GLMM	-	-	-	-	-	-	-	1	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-
28	Onozuka (2010)	Poisson GLM with DLM	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-
53	Paz (2009)	Poisson GLM	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
59	Philipsborn (2016)	Poisson GLM; Poisson GEE (meta-regression)	1	-	-	-	-	-	-	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
29	Phung (2015a IJB)	Poisson GLM with DLM	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-
54	Reyburn (2011)	SARIMA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-

#	Study (First Author, Year)	Model(s)	multiple model types (n=18)	OLS (n=2)	multiple linear (n=3)	logistic (n=8)	binomial (n=1)	log-linear (n=3)	GAM (n=12)	Poisson GLM (n=47)	Poisson GAM (n=8)	Poisson GEE (n=2)	NBRM (n=19)	NB GAM (n=2)	GLMM (n=7)	GAMM (n=1)	random effects (RE) (n=4)	GLS (n=1)	Bayesian Poisson (n=1)	BSTHM (n=2)	auto-regressive (AR) (n=1)	(S)ARIMA(X) (n=6)	DLM (n=13)	DLNM (n=18)	zero-inflated (n=1)	case-crossover (n=4)	meta-regression/-analysis (n=8)
30	Seidu (2013)	Poisson GLM	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
31	Singh (2001)	multiple linear regression; Poisson GLM	-	-	1	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
93	Song (2018)	Poisson GAM	-	-	-	-	-	-	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
77	Thindwa (2019)	Poisson GLM with DLNM	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-
32	Thompson (2015)	Poisson GLMM	-	-	-	-	-	-	-	1	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-
55	Trærup (2011)	NBRM	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
78	Wang (2018a EI)	NB GAM with DLNM	-	-	-	-	-	-	1	-	-	-	-	1	-	-	-	-	-	-	-	-	-	1	-	-	-
61	Wang (2018b STOTEN)	NBRM with DLNM; NB GAM with DLNM	1	-	-	-	-	-	1	-	-	-	1	1	-	-	-	-	-	-	-	-	-	1	-	-	-
33	H. Wang (2019)	Poisson GLM with DLNM	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-
34	Wang (2021)	Poisson GLM with DLNM (meta-analysis)	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	1
35	Wangdi & Clements (2017)	RE Bayesian Poisson model	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	1	-	-	-	-	-	-	-	-
44	White (2009)	Poisson GLM; case-crossover (time-stratified 2:1 matched)	1	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-
36	Wu (2014)	logistic regression	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
56	Wu (2018)	case-crossover (conditional logistic regression)	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-
37	Xu (2013)	Poisson GLM with DLNM	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-
38	Xu (2014)	Poisson GLM with DLNM	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-
94	Xu (2018)	BSTHM	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-
95	Yan (2017)	ARIMAX	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-
79	Yun (2016)	NBRM	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-

#	Study (First Author, Year)	Model(s)	multiple model types (n=18)	OLS (n=2)	multiple linear (n=3)	logistic (n=8)	binomial (n=1)	log-linear (n=3)	GAM (n=12)	Poisson GLM (n=47)	Poisson GAM (n=8)	Poisson GEE (n=2)	NBRM (n=19)	NB GAM (n=2)	GLMM (n=7)	GAMM (n=1)	random effects (RE) (n=4)	GLS (n=1)	Bayesian Poisson (n=1)	BSTHM (n=2)	auto-regressive (AR) (n=1)	(S)ARIMA(X) (n=6)	DLM (n=13)	DLNM (n=18)	zero-inflated (n=1)	case-crossover (n=4)	meta-regression/-analysis (n=8)
96	Zhang (2007)	SARIMA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-
80	Zhang (2008a IJB)	SARIMA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-
81	Zhang (2010)	Poisson GLM	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
97	Zhang (2021)	BSTHM; GeoDetector	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-
39	Zhou (2013)	Poisson GAM	-	-	-	-	-	-	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Model acronyms: ordinary least squares (OLS), generalized additive model (GAM), generalized linear model (GLM), negative binomial regression model (NBRM), negative binomial (NB), generalized linear mixed model (GLMM), generalized additive mixed model (GAMM), random effects (RE), generalized least squares (GLS) model, Bayesian space-time hierarchical models (BSTHM), (seasonal) autoregressive integrated moving average (with exogenous variables) ((S)ARIMA(X)), autoregressive (AR) model, distributed lag models (DLM), distributed lag non-linear models (DLNM)

TABLE 12. Example control terms for secular (long-term) and seasonal trends from four recent AGI-weather time series studies.

Example control terms for…		Control Option					
secular (long-term) trends	**seasonal trends**	**Time Stratified (1)**	**Periodic functions (2)**	**Splines (3)**	**Time UOA (day, week)**	**Model type(s)**	**Reference**
$\ldots + \mu_{year}$	$\ldots + \mu_{month} + s_{cubic}(v_t \mid df = 3; 6\ mo.)$	✓		✓	w	GAM	(Uejio et al., 2014)
$\ldots + s_{natural\ cubic}(v_{it} \mid df = 20)$	$\ldots + \beta_3 season_{it} \begin{cases} winter \\ spring \\ summer \\ fall \end{cases}$	✓		✓	w	negative binomial	(D. Lee et al., 2019)
$\ldots + s_{cubic}(v_t \mid df = 5/yr)$	$\ldots + \gamma_1 \sin\left(2 \cdot \pi \cdot \frac{t}{52}\right) + \gamma_2 \cos\left(2 \cdot \pi \cdot \frac{t}{52}\right)$		✓	✓	w	DLNM	(Chhetri et al., 2017)
same as control for seasonal trend	$\ldots + s_{natural}(v_t \mid df)$			✓	d	DLNM GAM	(Jagai et al., 2015b)
same as control for seasonal trend	$\ldots + s_{thin-plate}(v_t \mid df = 7(3-12)/yr)$			✓	d	DLNM GAM	(Tornevi et al., 2015)

Symbols: s *is a spline;* v *is an ordered discrete count of the time unit of observation (UOA) for a given study;* t *is the indexed time variable (day, week) matching the study's time unit of observation (UOA);* i *is the indexed spatial variable (e.g., county)*

2.8.1. Frequency indices of exceedance days for extreme temperature (ETT_{95}) and precipitation (EPT_{90})

Upperman and colleagues developed surrogate exposure metric for extreme heat events at the county-level based on climatology (1960–1989 reference period) using a frequency index based on exceeding specific percentile thresholds for daily maximum temperature (T_{max}) (Upperman et al., 2015). This calendar month and county-specific index was modified in two ways when applied in three additional weather-health studies in Maryland: (a) the index was defined for cumulative precipitation, in addition to maximum temperature, and (b) computed for each calendar day (instead of calendar month) using a 31-day window centered around each day (Jiang et al., 2015; Soneja, Jiang, Upperman, et al., 2016; Soneja, Jiang, Fisher, et al., 2016).

In these three studies, extreme temperature and precipitation are represented by the 'Extreme Temperature Threshold 95th percentile' (ETT_{95}) and 'Extreme Precipitation Threshold 90th percentile' (EPT_{90}) (Equation 1 and Equation 2 respectively). According to the definition in Schär *et al.* (Schär et al., 2016), these are frequency indices based on percentile thresholds because they are "percentile thresholds [that] are derived for some reference period, and the subsequent analysis then targets the frequency with which these thresholds are exceeded." Schär and colleagues (2016) note that frequency indices are officially recommended by the World Climate Research Program (WCRP) and World Meteorological Organization (WMO, 2009; X. Zhang et al., 2011), have been used in the following climate studies (Durman et al., 2001; Frei & Schär, 2001; Giorgi et al., 2014; Karl & Knight, 1998; Klein Tank & Können, 2003; Orlowsky & Seneviratne, 2012; Sillmann et al., 2013), and are a good alternative to wet-day percentile indices (along with all-day percentile indices and extreme value theory (EVT)).

As used by the three aforementioned studies (Jiang et al., 2015; Soneja, Jiang, Upperman, et al., 2016; Soneja, Jiang, Fisher, et al., 2016) and derived in Upperman *et al.* (2015), 'Extreme Temperature Threshold 95th percentile' (ETT_{95} or ETT_{jk-95}) is defined for each calendar month *k* and county *j* and is the sum of the number of 'exceedance days' ($I_{ijk} = 1$) per month, based on the binary indicator (I_{ijk}) (see Equation 1). An exceedance day occurs if the county-specific daily maximum temperature (TMAX or $T_{\mathrm{ijk-max}}$) is greater than the county (*j*) and day (*i*)-specific 95th percentile Tmax value ($\mathrm{T}_{\mathrm{ji\pm15-95}}$) in a 31-day window aroincludend day *i* for the reference period of 1960-1989. Similarly, the 'Extreme Precipitation Threshold 90th percentile' (EPT_{90} or EPT_{jk-90}) follows the same construction as ETT_{95}, but using total daily precipitation (PPT) and a thresh^ol^d of

90th percentile daily total precipitation ($\mathrm{PPT}_{\mathrm{ji}\pm15-90}$) (see Equation 2). The IPCC (Intergovernmental) Panel on Climate change accepts this 30-year period (1960-1989) as the standardized climate regime representation (Solomon et al., 2007).

Equation 1. Extreme Temperature Threshold 95th percentile (ETT_{95}) equation (Jiang et al., 2015; Soneja, Jiang, Upperman, et al., 2016; Soneja, Jiang, Fisher, et al., 2016; Upperman et al., 2015)

$$ETT_{jk-95} = \sum_{i} I_{ijk}$$

$$where\ I_{ijk} = \begin{cases} 1, if\ T_{\mathrm{ijk-max}} > \mathrm{T}_{\mathrm{ji}\pm15-95} \\ -, if\ \mathrm{T}_{\mathrm{ijk-max}} \leq \mathrm{T}_{\mathrm{ji}\pm15-95} \end{cases}$$

- where:
  - county *j*
  - calendar month *k*
  - day *i* of calendar month *k*
  - $ETT_{jk\text{-}95}$ is the total number of extreme heat events for county *j* in calendar month *k*
  - $T_{ijk\text{-}max}$ is the daily maximum temperature (T_{max}) in county *include* for day *i* of calendar month *k*;
  - $\mathrm{T}_{\mathrm{ji}\pm15-95}$ is the 95th percentile T_{max} value for county *j* in a 31-day window includeround day *i* (*i*+/-15 days) for the 30-year (1960–1989) period
  - I_{ijk} represents the indicator of whether or not $T_{ijk\text{-}max}$ is greater than T_{jk-95}

Equation 2. Extreme Precipitation Threshold 90th percentile (EPT_{90}) equation (Jiang et al., 2015b; Soneja, Jiang, Upperman, et al., 2016; Soneja, Jiang, Fisher, et al., 2016; Upperman et al., 2015)

$$EPT_{jk-90} = \sum_{i} I_{ijk}$$

$$where\ I_{ijk} = \begin{cases} 1, if\ PPT_{\mathrm{ijk-tot}} > \mathrm{PPT}_{\mathrm{ji}\pm15-90} \\ -, if\ \mathrm{PPT}_{\mathrm{ijk-tot}} \leq \mathrm{PPT}_{\mathrm{ji}\pm15-90} \end{cases}$$

- where:
  - county *j*
  - calendar month *k*
  - day *i* of calendar month *k*
  - $EPT_{jk\text{-}90}$ is the total number of extreme heat events for county *j* in calendar month *k*
  - $PPT_{ijk\text{-}tot}$ is the daily total precipitation (*PPT*) in county *j* for day *i* of calendar month *k*;
  - $PPT_{ji\pm15,k-95}$ is the 90th percentile total precipitation *PPT* value for county *j* in a 31-day window around day *i* (*i*+/-15 days) for the 30-year (1960–1989) period
  - I_{ijk} represents the indicator of whether or not $PPT_{ijk\text{-}tot}$ is greater than PPT_{jk-95}

3. ACUTE GASTROINTESTINAL ILLNESS AND INFLUENCE OF RAINFALL EXPOSURE MEASURES IN NORTH CAROLINA, USA (AIM 2)

3.1. Introduction

The intensity and frequency of extreme weather events—heavy and extreme precipitation events in particular— increased over the 20th century and are projected to further increase during the 21st century in many regions of the United States (Easterling et al., 2017; Lall et al., 2018; Prein et al., 2017). As a result, changes in microbial transmission and contamination are projected to exacerbate morbidity and mortality due to waterborne acute gastrointestinal (AGI) diseases, which are sensitive to temperature and hydroclimatic drivers such as water availability, rainfall, flooding, and drought (J. N. S. Eisenberg et al., 2007). Current and future extreme precipitation events may compound the challenges of aging infrastructure (Lall et al., 2018). For example, heavy rainfall can lead to sewer overflows and overwhelmed septic systems have been respectively associated with increased acute gastrointestinal illness (Jagai et al., 2015, 2017; Miller et al., 2022) and fecal contamination in nearby private wells (Murphy et al., 2020).

The relationships between precipitation and AGI are complex and they vary by location; studies conflict as to whether rainfall increases or decreases rates of diarrhea (Kraay et al., 2020). The use of multiple, non-standardized precipitation exposure measures (e.g., continuous measures of rainfall, heavy or extreme rainfall, or antecedent rainfall conditions) is a source of additional variation between studies on diarrhea and weather that makes study comparison and meta-analyses more challenging (Kraay et al., 2020; K. Levy et al., 2016). In their systematic review, Kraay and colleagues (2020) report that in a few studies extreme rainfall appears to be modified by antecedent wet and/or dry conditions, though effect measure modification varies amongst the four studies and needs further study. They present compelling evidence from rainfall, extreme rainfall, season, flood, and drought studies to support concentration-dilution dynamics, and conclude by calling for the use of standard, clearly defined exposure measures for rainfall to improve study comparability and our understanding of the relationship between rainfall and diarrheal illness (Kraay et al., 2020). This study assesses the sensitivity of the association between AGI and weather in North Carolina to multiple measures of precipitation.

The southeastern US is projected to experience major climate change impacts from extreme precipitation events, hurricanes, flooding, and rising sea levels over the next century (Lall et al., 2018). North Carolina is uniquely positioned to model the health effects of hydroclimatic variability as the state is heterogeneous both geographically (Raisz, 1940) and climatologically (State Climate Office of North Carolina, n.d.) with large, growing, and diverse urban and rural populations. A recent study attributed 29,000 ED visits and $40 million in associated costs annually to microbial contamination of drinking water in the state (DeFelice et al., 2016). Statewide studies of all-cause or pathogen-specific AGI and weather-related events have been conducted in Georgia, Massachusetts, New Jersey, and North Carolina for precipitation (NJ: Gleason & Fagliano, 2017; MA: Jagai et al., 2015; GA: D. Lee et al., 2019), flooding (NC: Quist, Fliss, et al., 2022; MA: Wade et al., 2014), and combined or sanitary sewer overflows (MA: Jagai et al., 2017; GA: Miller et al., 2022). Bivariate statistical analyses at the monthly, county level (Hartley, 2016) found increased rates of gastrointestinal illness after periods of heavy rainfall, but a more rigorous, multivariable time series regression or case-crossover study of the relationship between acute gastrointestinal illness and weather has not been conducted for North Carolina.

This study responds to the call of Kraay *et al.* (2020) by systematically comparing different definitions of multiple precipitation measures (absolute, extreme, and antecedent precipitation) in the same study area using high resolution AGI outcome and weather data (daily, ZIP code) to develop time series models for North Carolina. We ask two questions in this study. What is the relationship between AGI) emergency department (ED) rates and precipitation in the state of North Carolina? How sensitive is the association between AGI ED rates and precipitation to different precipitation exposure definitions?

3.2. The concentration-dilution hypothesis and precipitation exposure measures

In a systematic review and meta-analysis of hydrometeorology and diarrheal illness studies, Kraay and colleagues (2020) reviewed and calculated pooled estimates for rain, extreme rain, season, flood, and drought. Kraay and colleagues evaluated the reviewed these studies with respect to an expanded framework of the concentration-dilution hypothesis that includes flooding, drought, rainfall in arid climates, and season, in addition to the original extreme precipitation. According to the concentration-dilution hypothesis, the effect of rainfall or extreme rainfall depends on the antecedent or background rainfall conditions, a concept to explain conflicting findings about the direction of the relationship between rainfall and diarrheal illness through flushing, runoff, concentration, and dilution effects, for which time-dependent mechanisms are described by Levy and colleagues (2009) and Moors and colleagues (2013), summarized within systematic review of temperature, heavy rainfall, flooding, and drought by Levy and colleagues (2016), and systematically reviewed by Kraay and colleagues (2020). Dry periods may allow pathogen accumulation in the environment, creating a concentration effect that increases diarrheal risk when an extreme precipitation event occurs and flushes pathogens into the environment and surface waters. The first flush or flushing effect may also increase diarrheal risk and occurs initial volumes of rain in urban areas produce runoff or stormwater discharge with higher pollutant concentrations (Bach et al., 2010; Bertrand-Krajewski et al., 1998), and a seasonal first flush is when initial storms in a wet season have higher pollutant concentrations after pollutants build up during the prior dry season (H.

Lee et al., 2004). Similarly, the runoff effect similarly flushes and resuspends pathogens into surface waters after a heavy rain event and may increase diarrheal risk, but may occur throughout the year, especially if there is a continual source of microbial contamination or fecal material (K. Levy, Hubbard, Nelson, et al. 2009). Conversely, extreme rain events following wet periods may create conditions for the dilution of pathogens in surface waters and decrease in diarrheal risk (protective effect).

There are three general types of rainfall or precipitation exposure definitions: rain (a.k.a. absolute precipitation), heavy or extreme precipitation, and antecedent precipitation characterizing wet and dry periods prior to a rain event.

The first type of precipitation exposures variables, which we refer to as absolute precipitation to distinguish it from the relative measures of extreme precipitation, are continuous measures of rainfall, such as the cumulative/total or average amount of rainfall over given period of time (e.g., daily, weekly monthly, over prior 7 d, over prior 15 d) (Kraay et al., 2020). In the systematic review by Kraay and colleagues, 50 articles in the qualitative analysis and 15 in the quantitative meta-analysis included rain (absolute rainfall) as a climate exposure. While they found no linear association between rain and diarrhea in the meta-analysis, 5 studies had nonlinear associations where diarrheal risk increased at high and low rainfall levels (U-shaped) (Fang et al. 2019; Dunn and Johnson 2018; Ikeda et al. 2019), moderate rainfall (Chowdhury et al. 2018), and high levels only (Uejio et al. 2014). To assess nonlinearities, the authors recommended using continuous rainfall exposure variables instead of categorizing rainfall *a priori*.

The second type of precipitation variable is heavy or extreme precipitation (e.g., 80th, 90th, 95th, and 99th percentile), which are relative measures that use a precipitation percentile index to create binary or categorial indicator variables for extreme and non-extreme events over a period of time. Precipitation percentile indices are popular to characterize heavy or extreme precipitation in climate-diarrheal studies (Carlton et al., 2014; Curriero et al., 2001; K. M. Thomas et al., 2006), but their exposure measure definitions vary and it is not well-known whether or how sensitive estimates may be to the precipitation definitions. While heavy or extreme precipitation variables may provide a useful location-based variable, the relative nature of these percentile-based measures that are based on local climatology limits that their generalizability and can makes it challenging to compare across studies. Furthermore, extreme precipitation measures look similar on the surface, but differences emerge when the exposure measures are examined more closely. While it is more common for AGI and precipitation studies to conduct sensitivity analyses on the percentile thresholds defining the extreme (i.e., 80th, 90th, 95th, 99th) or the length of exposure lags, extreme precipitation definitions are not standardized and the details of their definitions and the precipitation cutpoints (mm) corresponding to a given threshold are inconsistently reported. We've observed the following differences in extreme precipitation definitions: inclusion or exclusion of zero-precipitation or dry days, which respectively define all-day or wet-day precipitation percentile indices; different definitions of heavy or extreme indicator variables (total, cumulative, or average values) over various periods of temporal aggregation (e.g., daily, weekly, monthly); the precipitation thresholds that define a wet day (e.g., > 0 or 1 mm/d or 0.1 mm/h); inconsistencies with the length of the precipitation reference periods over which the precipitation distribution and percentile cutpoints (mm) are calculated (e.g., length of study period or 15, 20, or 30 years); and the spatial extents (e.g., specific to a given weather station, ZIP code, county or defined over multiple weather stations or spatial areas) that are included calculating the precipitation distribution and cutpoints. When sample sizes are low and data must be temporally and/or spatially aggregated to achieve sufficient power, researchers are also faced with tradeoffs defining weather exposures over time and space to match the outcome unit analysis and still be resolved enough to capture the relationship between the outcome and exposure.

For example, some climate studies suggest that precipitation percentile indices (all-day percentiles (≥ 0 mm or 1 mm/d or 0.1 mm/h), wet-day percentiles (> 0 or 1 mm/d or 0.1 mm/h), and frequency indices based on the exceedance of a percentile threshold) used to characterize heavy or extreme precipitation may lead to different results and wet-day percentiles may be sensitive to the frequency and magnitude of rainfall events. In addition to being used in environmental epidemiology studies, precipitation indices (e.g., percentile indices) are also frequently used in climate change studies to assess the trends and projections of heavy precipitation events and could be a valuable source of information to dig deeper into the how precipitation and temperature variables can be defined for predictors at different scales. Multiple climate studies more rigorously (and mathematically) defined various precipitation indices and investigated their performance and sensitivities across different conditions. Schär and colleagues (2016), for example, define and assess the comparability and robustness of three types of common heavy precipitation percentile indices in climate applications: all-day percentile indices (includes zero-precipitation days), wet-day percentile indices, and frequency indices (based on the frequency of exceedance of threshold). The authors found that wet-day precipitation indices are more sensitive to frequency and magnitude of wet days and recommended the use of all-day percentile indices or frequency-based precipitation indices as an alternative to wet-day percentile indices. Schär and colleagues (2016) also note that frequency indices are officially recommended by the World Climate Research Program (WCRP) and World Meteorological Organization (WMO, 2009; X. Zhang et al., 2011), have been used in the following climate studies (Durman et al., 2001; Frei & Schär, 2001; Giorgi et al., 2014; Karl & Knight, 1998; Klein Tank & Können, 2003; Orlowsky & Seneviratne, 2012; Sillmann et al., 2013). An equivalent metric for temperature would be degree-days, commonly used in climate change economics and environmental health studies. A good example of temperature and precipitation frequency exceedance indices developed for environmental health studies, which address spatiotemporal aggregation tradeoffs by defining a categorical exposure variable at the county-month level based on more resolved daily, county-level data, are the Extreme Temperature Threshold 95th percentile (ETT_{95}) and Extreme Precipitation Threshold 90th percentile (EPT_{90}) metrics. Upperman and colleagues (2015) developed ETT_{95} as surrogate exposure metric for extreme heat events at the county-level based on climatology (1960–1989 reference period) using a frequency index based on exceeding specific percentile thresholds for daily maximum temperature (T_{max}). It was subsequently expanded to include precipitation (EPT_{90}) in applications of salmonellosis (Jiang et al., 2015), campylobacteriosis (Soneja, Jiang, Upperman, et al., 2016), and asthma (Soneja, Jiang, Fisher, et al., 2016) at the county-month scale in Maryland (see Supplementary Equation 2 for the EPT_{90} definition).

The third type of precipitation variable—and one of the most interesting precipitation measures in development—is antecedent precipitation, which are exposure variables that try to test the concentration-dilution hypothesis (Kraay et al., 2020; K. Levy, Hubbard, Nelson, et al., 2009; Moors et al., 2013) or first flush theory (Bach et al., 2010; Bertrand-Krajewski et al., 1998) by defining wet or dry periods to estimating the effect of heavy precipitation on antecedent precipitation conditions. There are at least eight recent climate-diarrhea studies that use various measures for antecedent precipitation (K. F. Bush et al., 2014b; Carlton et al., 2014; Chhetri et al., 2017; Graydon et al., 2022; D. Lee et al., 2019; Mertens et al., 2019; Tornevi et al., 2013, 2015), four of which were included in the systematic review by Kraay and colleagues. Developing and testing better, broadly applicable, and, ultimately, more standardized measures of antecedent precipitation exposures for different levels of spatiotemporal aggregation and extent is an important new area of research.

3.3. Materials and methods

Using high resolution outcome and weather data, we employed time series analysis to estimate the statewide association between precipitation and AGI ED rates in NC at the daily, ZIP code level (2018-2015) with quasi-Poisson generalized linear models (GLM) and distributed lag models (DLM). We empirically tested various daily precipitation measures: absolute precipitation (daily total precipitation); four definitions of 90^{th}, 95^{th}, and 99^{th} percentile extreme precipitation; and two measures of antecedent precipitation to explore the effect of different climate exposure definitions on the relationship between AGI emergency department (ED) rates and precipitation.

3.3.1. Study design and population

We conducted a retrospective time series analysis of meteorological variables (air temperature and daily rainfall) and emergency department (ED) visits related to any acute gastrointestinal illness (AGI) diagnosis in North Carolina (NC) residents from January 1, 2008 to September 30, 2015. All-cause AGI cases were defined as ED visits by NC residents having a non-post office box billing ZIP code within NC matched with ZIP code polygons from 2013 (ESRI, Redlands, CA, USA; 2013) who were assigned an AGI-related ICD-9-CM diagnosis code during the study period. The study unit-of-analysis (UOA) was defined at the daily, ZIP code level because it was the finest temporal and spatial resolution of outcome data available for ED visits.

3.3.2. Outcome data and definitions

Combined with a ZIP code-level population offset, emergency department (ED) visits for AGI in NC were used as an indicator for AGI incidence, with the recognition that only a fraction of AGI cases are treated in an emergency department (Bylund et al., 2017). Daily ED visit data was obtained from NC DETECT, North Carolina's statewide syndromic surveillance system (Carolina Center for Health Informatics, University of North Carolina at Chapel Hill, 2010; Hakenwerth et al., 2009; Lippmann et al., 2013; Waller et al., 2011). We assume ED patients contracted AGI in their zip code where they reside.

The study period was defined as January 1, 2008 to September 30, 2015. By 2008, 99.5% of all NC emergency department visits were estimated to be captured by NC DETECT. The study end date was defined as September 30, 2015 to avoid the shift from ICD-9-CM to ICD-10-CM as the diagnostic standard on October 1, 2015 which may have impacted disease trends and/or case definitions. We used NC DETECT patient information on ZIP code and state of residence, insurance type, the date emergency department visit, and up to 11 discharge diagnoses coded to the 9th Revision of the International Classification of Diseases–Clinical Modification (ICD-9-CM) discharge diagnosis codes (USDHHS et al., 2009)[9] for AGI: infectious GI illness (001.xx to 009.xx), non-infectious GI illness (558.9), diarrhea (not otherwise specified) or nausea, vomiting, and diarrhea (787.91), and nausea and/or vomiting (787.01-787.03), following recent studies (DeFelice, 2014; DeFelice et al., 2015; Tinker et al., 2009, 2010). Refer to Appendix TABLE 25 for ICD-9-CM descriptions and full descriptions of ICD-9-CM diagnosis codes and a comparison of derived case definitions from prior studies in Appendix TABLE 26.

3.3.3. Meteorological exposure data and definitions

PRISM gridded weather dataset

Gridded climate datasets (GCD) can help address the limitations of weather station data, particularly those related to inhomogeneous spatial distribution (Auffhammer et al., 2013; Mourtzinis et al., 2017; Spangler et al., 2019). GCDs use weather station data and additional satellite-derived data to estimate high resolution data for multiple meteorological variables over large spatial areas, which is useful to cover larger populations in environmental epidemiology. We used the daily gridded AN81d gridded climate dataset from PRISM (Parameter-elevation Regressions on Independent Slopes Model; Daly et al., 2008; PRISM Climate Group, 2004), aggregated to the ZIP code level to define meteorological exposures from daily total precipitation and daily mean temperature for this study.

---

[9] All available data elements are listed on the NC DETECT website (https://ncdetect.org/data-elements/), but the data is subsetted based on the data use agreement (DUA) between the researcher(s) and NC DETECT.

The PRISM AN81d dataset is a high resolution (2.5-arcmin or 4x4-km) grid of daily meteorological data for the conterminous United States (CONUS) available in a stable dataset from January 1981 to 6 months prior to present. PRISM products incorporate weighted weather station point data, a digital elevation model (DEM), topographic and other geophysical features using statistical interpolation (Daly et al., 1994, 2008). A PRISM day is defined as the 24-hour period prior to noon UTC, but was assumed to match the midnight-to-midnight days defined by NC DETECT. The PRISM AN81d dataset include estimates of seven climate elements: total precipitation (PPT), mean, maximum and minimum temperature (T_{mean}, T_{max}, T_{min}), mean dewpoint temperature (Td_{mean}), and maximum and minimum vapor pressure deficit (Vpd_{max}, Vpd_{min}). Relative humidity (RH), absolute humidity (AH), and heat index (HI) can also be derived from the reported variables (Spangler et al., 2019).

Data aggregation and transformation for weather variables

Area-weighted spatial averaging (Dell et al., 2014) was used to create a balanced daily, ZIP code-level time series dataset of weather exposures (PPT, T_{mean}, T_{max}, T_{min}) by aggregating the daily 4x4-km grid cell data to 737 ZIP code polygons from 2013 (ESRI, Redlands, CA, USA; 2013) in ArcGIS 10.5.1 by limiting the spatial extent to North Carolina, uniformly downscaling the 4x4-km gridded data to 1x1-km grids, and calculating daily spatial statistics (spatial mean, median, maximum, minimum) across all 1x1-km grid centroids within a given ZIP code polygon.

Precipitation measures

The ZIP code spatial mean of daily total precipitation (PPT) was used to create 3 types of measures for modeling the association between AGI ED rates and precipitation: *Absolute PPT*, *Extreme PPT*, and *Antecedent PPT*.

1. **Absolute PPT:** continuous (a.k.a. absolute) measure of daily total precipitation (spatial mean of grids within each ZIP code on day *t* and ZIP code *i*).

2. **Extreme PPT:** For extreme precipitation, we considered the threshold of the percentile (90^{th}, 95^{th}, and 99^{th}), the spatial reference (ZIP code and statewide), and inclusion or exclusion of zero-precipitation days. Extreme precipitation percentile indices were calculated two ways: as "all-day" indices using all available data (PPT ≥ 0 mm), and as "wet-day" indices using only those days with greater than 0 mm of daily total precipitation (PPT > 0m). Thus we considered a total of 12 different indices for extreme PPT.
3. **Antecedent precipitation:** We tested two measures of antecedent precipitation measures that use categorical indicator variables for prior 8-week wet-try tertiles and consecutive wet-dry days based on recent studies.
    a. *Wet-dry tertiles* (Carlton et al., 2014; D. Lee et al., 2019): 8-week (56-day rolling sums) cumulative precipitation sums prior to day *t* were calculated for each day and then used to calculate precipitation cutpoints for the 33^{rd} and 67^{th} percentiles relative to spatial reference area (statewide, county-specific, ZIP code-specific), and classify 8-week precipitation tertiles into a categorical indicator variable (wet , moderate, dry periods). The daily antecedent precipitation variable definition of 8-week wet-dry tertiles prior to day *t* were modified from the weekly antecedent precipitation definition (8-week wet-dry tertiles prior to week *t*) used by Carlton and colleagues (2014) (single value for the spatial area; Ecuador) and Lee and colleagues (2019) (county-specific; Georgia).
    b. *Wet-dry days* (Tornevi et al., 2013, 2015): A categorical indicator variable was calculated for the length of consecutive wet (+) and dry (-) days prior to day *t,* modified from Tornevi and colleagues. Consecutive wet and dry days were adapted into categories for North Carolina into 1 wet/dry (reference), 2-3 wet, 2-3 dry, 4-6 wet, 4-6 dry, 7+ wet, 7+ dry. In a sensitivity analysis, a continuous measure of the

we-dry day antecedent precipitation was defined as the number of consecutive wet (+) and dry (-) days and used in an 8-day distributed lag model (DLM).

3.3.4. Covariates

All time series models included adjustments for covariates and potential confounding variables. We controlled for daily same-day ambient temperature and the following controls for confounding common in time series analysis. Short-term time effects were controlled by day-of-week (DOW), which provided a better fit than weekend/weekday, and U.S. federal holidays ($Holidays_t$) as defined by the New York Stock Exchange (NYSE) using the holidayNYSE() function from the timeDate package in R for 2008-2015: New Year's Day, Martin Luther King, Jr. Day, Washington's Birthday, Good Friday, Memorial Day, Independence Day, Labor Day, Thanksgiving Day, and Christmas Day. Long-term trends and seasonality were adjusted using a natural spline of day-of-year (DOY_t) with six degrees of freedom interacted with an indicator for year ($ns(DOY_t, df = 6)/Year_l$) following Thomas and colleagues (2021). In exploratory analyses, the time spline and year interaction terms performed better than a single natural spline with 55 knots [7 x number of calendar years (8) – 1], which is a common control for daily mortality studies (Bhaskaran et al., 2013; Dominici, Samet, et al., 2000). We used a fixed effect indicator variable for NC counties ($County_i$) to control for geographic variation in the outcome. For a seasonal sensitivity analysis, some models also controlled for season using a 4-level categorical variable for season ($Season_t$).

3.3.5. Statistical analyses

To estimate the statewide association between all-cause AGI ED rates and precipitation for North Carolina, we used quasi-Poisson generalized linear models (GLM) to account for overdispersion and county-level fixed effects time series analyses at the daily, ZIP code level, as specified in the general model (Equation 3). We applied a log offset of annual ZIP code population ($Population_{il}$) estimated from 5-year American Community Survey (ACS) block group population data and centroid locations aggregated to 2013 ZIP codes (Environmental Systems Research Institute (ESRI), 2013). We tested the sensitivity of the relationship between all-cause AGI ED rates to precipitation measure definitions (PPT_{it}) by using 4 sets of models modifying the general equation (Equation 3). Model 1 defined the exposure variable as *Absolute PPT* (Equation 4), Model 2 as *Extreme PPT* (Equation 5), Model 3 as *Antecedent PPT* (Equation 6), and Model 4 as the interaction between extreme and antecedent precipitation, *Extreme PPT x Antecedent PPT* (Equation 7). Distributed lag models (DLM) (Gasparrini, 2011; Gasparrini et al., 2010) have been used in some time series studies of weather and diarrheal disease or acute gastrointestinal illness to estimate the cumulative association of weather exposures over multiple lags (e.g., Hall et al., 2011; Jagai et al., 2015; Phung et al., 2015). In models using absolute precipitation (Model 1) or extreme precipitation (Model 2) to estimate the cumulative association between precipitation and AGI ED rate over the entire lag period, precipitation was specified as linear DLM term using a 3rd degree polynomial and precipitation lags (lag) of l days were evaluated over 0-7 days. Distributed lag models (DLM) were conducted in R version 4.2.0 using the package "dlnm" (Gasparrini, 2011).

Equation 3. General model

$$\log(E[AGI\ ED\ visits_{it}]) \sim \beta_0 + \beta_1 PPT_{it} + \beta_2 TEMP_{it} + \beta_3\ County_i + \left(\beta_4 Season_t\right) + \beta_5 DOW_t$$
$$+ \beta_6 Holidays_t + \beta_7\ \mathrm{ns}(DOY_t, \mathrm{df} = 6)/Year_t + log\left(Population_{it}\right)$$

Where:

$AGI\ ED\ visits_{it}$	time series of daily emergency department visits for all-cause AGI on day *t* in ZIP code *i*;
PPT_{it}	time series of the spatial mean of daily total precipitation (PPT) on day *t* in ZIP code *i*, as defined by model and precipitation measure;
$TEMP_{it}$	time series of daily mean ambient temperature on day *t* in ZIP code *i*;
$County_i$	indicator variable for the NC county enclosing the maximum spatial area for ZIP code *i*;
$Season_t$	[seasonal sensitivity analysis in some models] categorical indicator variable to control for 4 meteorological seasons by month: spring (March-May), summer (June-August), autumn (September-November;

	reference category), and winter (December-February) (Curriero et al., 2001; Jagai et al., 2015; Nichols et al., 2009; Upperman et al., 2015);
DOW_t	categorical indicator variable (reference category: Monday) to control for day-of-week (DOW) on day *t*;
$Holidays_t$	binary indicator variable (0/1) to control for federal holidays on day *t*, as defined by New York Stock Exchange (NYSE) holidays (source: R package *timeDate* using the *holidaysNYSE* function based on https://www.nyse.com/markets/hours-calendars);
$\text{ns}(DOY_t, \text{df} = 6)/Year_t$	natural cubic spline (n*s*) function for calendar day-of-year (DOY*t*) with 6 degrees of freedom (*df*), interacted with year to control for long-term trends and seasonality (N. Thomas et al., 2021);
$log(Population_{it})$	offset of log of annual population for ZIP code *i*, which allows the rate of ED visits to be calculated.

In Model 1 (Equation 4: *Absolute PPT*), absolute precipitation was defined as the spatial mean of total precipitation (mm) per day *t* in ZIP code *i* ($AbsolutePPT_{i,t-l}$), specified as a DLM over lags of 0-7 days. The cumulative association between daily AGI ED rates and absolute precipitation was evaluated at 1, 10, 20, 30, 40, 60, and 80 mm of daily total precipitation (ref = 0mm).

Equation 4. Model 1: Absolute Precipitation (7-day DLM of daily total precipitation)

$$log[E(AGI\ ED\ visits_{it})]$$

$$= (\beta_0 + b_i) + \sum_{l=0}^{lag=7} \beta_1\, AbsolutePPT_{i,t-l} + \beta_2 TEMP_{it} + \beta_3\, County_i + (\beta_4 Season_t)$$

$$+ \beta_5 DOW_t + \beta_6 Holidays_t + \beta_7\, \text{ns}(DOY_t, \text{df} = 6)/Year_t + log(Population_{it})$$

In Model 2 (Equation 5: *Extreme PPT*), to estimate the cumulative association between the daily rate of AGI ED visits and X$^{\text{th}}$ percentile extreme precipitation events (daily total precipitation greater than or equal to the 90$^{\text{th}}$, 95$^{\text{th}}$, 99$^{\text{th}}$ percentiles), we specified extreme precipitation as DLM term over lags of 0-7 days and evaluated four different measures of extreme precipitation indices varied by index type and spatial reference area ($ExtremePPT_{t-l}$): statewide all-day, ZIP code-specific all-day, statewide wet-day, and ZIP code-specific wet day.

Equation 5. Model 2: Extreme Precipitation (7-day DLM of 90th, 95th, 99th percentiles: all-day, wet day; statewide, ZIP code specific)

$$log[E(AGI\ ED\ visits_{it})]$$

$$= (\beta_0 + b_i) + \sum_{l=0}^{lag=7} \beta_1\ ExtremePPT_{t-l} + \beta_2 TEMP_{it} + \beta_3\ County_i + (\beta_4 Season_t)$$

$$+ \beta_5 DOW_t + \beta_6 Holidays_t + \beta_7\ \mathrm{ns}(DOY_t, \mathrm{df} = 6)/Year_t + log(Population_{it})$$

In Model 3 (Equation 6: *Antecedent PPT*), we estimated the association between daily AGI ED rates and antecedent precipitation using two different measures: consecutive wet-dry days (categories: 1 wet/dry (reference), 2-3 wet, 2-3 dry, 4-6 wet, 4-6 dry, 7+ wet, 7+ dry) (Tornevi et al., 2013, 2015) and 8-week wet-dry tertiles (categories: wet, moderate (reference), dry) (Carlton et al., 2014; D. Lee et al., 2019).

Equation 6. Model 3: Antecedent Precipitation (consecutive wet-dry days; 8-week wet-dry tertiles: statewide, county-specific, ZIP code-specific)

$$log[E(AGI\ ED\ visits_{it})]$$

$$= (\beta_0 + b_i) + +\beta_1 Antecedent_{it} + \beta_2 TEMP_{it} + \beta_3\ County_i + (\beta_4 Season_t) + \beta_5 DOW_t$$

$$+ \beta_6 Holidays_t + \beta_7\ \mathrm{ns}(DOY_t, \mathrm{df} = 6)/Year_t + log(Population_{it})$$

In Model 4 (Equation 7: *Extreme PPT x Antecedent PPT*), we used an interaction term for effect measure modification (EMM) to test whether antecedent precipitation modified the effect of extreme precipitation on AGI ED rates. Antecedent precipitation was defined as 8-week wet-dry tertiles (statewide and ZIP code-specific). It is not possible to interact a DLM term with another predictor, so we defined extreme precipitation as all-day 95th percentile (statewide and ZIP code-specific) on day *t* (lag = 0) and matched the spatial reference areas (state, ZIP code) between interacted extreme and antecedent precipitation terms. Effect measure modification can also be explored through stratification, but it would be complicated to stratify extreme precipitation days by prior 8-week dry, moderate, and wet periods because temporal stratification (unlike spatial stratification) would create an unbalanced time series dataset.

Equation 7. Model 4: Extreme (95th pct, lag = 0 d: statewide, ZIP code-specific) X Antecedent Precipitation (8-week wet-dry tertiles: statewide, ZIP code-specific)

$$log[E(AGI\ ED\ visits_{it})]$$

$$= (\beta_0 + b_i) + +\beta_1 Antecedent_{it} + \beta_2 TEMP_{it} + \beta_3\ County_i + (\beta_4 Season_t) + \beta_5 DOW_t$$

$$+ \beta_6 Holidays_t + \beta_7\ \mathrm{ns}(DOY_t, \mathrm{df} = 6)/Year_t + log(Population_{it})$$

3.3.6. Sensitivity analyses

For studies using all-cause AGI as an outcome, stratifying models by season or including seasonal indicator variables isa useful method to investigate weather patterns that may vary due to changes in the etiology underlying AGI throughout the year. For this study, the seasonal sub-analysis used used the common meteorological definition of four Northern Hemisphere seasons by month: spring (March, April, May), summer (June, July, August), fall (September, October, November; reference), and winter (December, January, February).

This study received an exemption from the University of North Carolina Institutional Review Board (Study #: 15-1158).

3.4. Results

From January 1, 2008 to September 30, 2015, there were 2,776,870 total all-cause AGI ED visits identified in the statewide North Carolina study population (TABLE 15). The frequency of AGI ED visits is described by year, season, patient characteristics (age group, sex, health insurance type), and physiographic region (Mountains, Piedmont, Coast). Descriptive statistics for alternative AGI or pathogen-specific case definitions, subsets of all-cause AGI, are included in TABLE 15, but were not modeled in this study: all-case AGI without nausea and vomiting (no NV) (1,079,955), bacterial AGI (46,705), viral AGI (79,379), protozoan AGI (1,157), all *Escherichia coli* or *E. coli*) (279), *Clostridium difficile* or *C. difficile* (37,176), all cholera (120). Annual all-cause AGI ED visits generally increased over time and ranged between 304,567 to 391,633 (2008-2014), with 281,104 AGI-related ED visits from January to September 2015. Seasonal patterns varied by AGI case definition, with the highest prevalence for all-cause AGI in winter (27%; 29%) and spring (27%; 29%) and lowest in fall (22%; 20%). Though the majority of all-cause AGI ED visits were from patients between 18-64 years old, 22% of patients were children and youth from 0-17 years old (12% 0-4 y; 5.8% 5-11 y; 4.2% 12-17 y) and 14% amongst patients 65 years or older. Most AGI-related ED visits were by women (all-cause AGI: 64% F; 36% M) for all case definitions except protozoan AGI, with and without abscesses (32% F; 68% M). Most patients with reported all-cause AGI-related had public health insurance (49%), followed by private health insurance (24%), self-pay or uninsured (21%), and other/unknown (5.9%), with similar patterns for all case definitions. The Piedmont region in central North Carolina, the most populous region of containing NC's the 5 largest cities, had the highest relative frequencies ED visits for all-cause AGI (62%), no-NV all-cause AGI (62%), bacterial AGI (64%), *C. difficile* (66%), and viral AGI (62%). However, the Mountains had the highest prevalence of cholera (72%) and protozoan AGI (54%), while the Piedmont and Coastal regions had number of ED visits due to *E. coli* (47% each).

The distribution of all-cause AGI ED visits and meteorological exposure variables (mean ambient temperature (*Tmean*), and absolute, extreme, and antecedent precipitation (*PPT*)) at the ZIP code-day level is shown in TABLE 16 (additional descriptive statistics for minimum and maximum temperature and absolute precipitation are available in Supplementary TABLE 18). From 2008-2015, 1,051,347 ZIP code-days (50.7%) in North Carolina had at least one all-cause AGI-related ED visit, with a daily average of 1.3 ± 2.2 SD and maximum of 31 daily ED visits per ZIP code. The average daily total precipitation (*Absolute PPT*) per ZIP code was 3.4 mm ± 8.9 SD of daily total precipitation for all days and 7.7 mm ± 9.9 SD amongst rainy days (*PPT* >0mm).

Supplementary FIGURE 2 shows the histogram of ZIP code-specific *Extreme PPT* cutpoints (mm) for daily total precipitation between all-day (left) and dry-day (right) percentiles. The statewide cutpoints (mm) are indicated by the red vertical line. As expected, wet-day percentiles correspond to higher precipitation cutpoints than all-day percentiles. For example, the statewide 95th all-day percentile is at 19.6 mm, which is closest to the 85th wet-day percentile (20.1 mm). The 99th percentiles for the statewide all-day and wet-day precipitation indices are respectively 43 and 63 mm. The mean and standard deviation of ZIP-code specific cutpoints increases for both all-day and wet-day percentiles as percentiles increase. The precipitation cutpoints also vary spatially across the state. The greatest positive deviations are on the coast and the southernmost areas of the Blue Ridge Mountains near the border with South Carolina (Supplementary FIGURE 5). The average number of extreme precipitation days per ZIP code by percentile amongst the all-day or wet-day indices was similar, but the standard deviation was greater for statewide than ZIP code-specific percentiles. Out of 2830 days in the study period, there were approximately 286 (90th), 139-140 (95th), and 27 (99th) extreme all-day precipitation days per ZIP code compared to 120 (90th), 59 (95th), and 12 (99th) extreme wet-day precipitation days per ZIP code from 2008-2015 (TABLE 16).

We used two measures of antecedent precipitation (*Antecedent PPT*): 8-week wet-dry tertiles (Carlton et al., 2014b; D. Lee et al., 2019) and cumulative wet-dry days (Tornevi et al., 2013, 2015). Based on a wet day threshold of greater than 0 mm of precipitation, North Carolina had approximately 1.15 million dry days and 929,416 wet days during the study period (TABLE 16). Using Tornevi and colleagues' (2013, 2015) consecutive wet-day measure, the majority of ZIP code-days were classified as only 1 consecutive wet or dry day (over 2.08 million) and the maximum consecutive dry (wet) period was 10 (14) days (TABLE 16; Supplementary TABLE 19 lists counts of antecedent wet and dry days used to define the categorical indicator variable for Antecedent PPT wet-dry days). Using a daily rolling rum of 8-week prior cumulative precipitation to define *Antecedent*

PPT wet-dry tertiles (dry, moderate, wet) (Carlton et al., 2014; Lee et al., 2019), the number of ZIP code-days classified as having followed an 8-week dry, moderate, or wet period were similar between the statewide, county-specific, and ZIP code-specific measures with the following ranges by tertile: 617,631 to 620,841 (dry); 735,543 to 737,156 (moderate); and 665,207 to 666,804 (dry). Cumulative precipitation cutpoints corresponding to 8-week tertiles by state, county, and ZIP code are shown in Supplementary TABLE 19.

FIGURE 1 shows the rate ratios (RR) and 95% confidence interval values for the statewide association between AGI ED rates and *Absolute PPT* (Model 1: Panel A), *Extreme PPT* (Model 2: Panel B), *Antecedent PPT consecutive wet-dry days* (Model 3: Panel C), and *8-week Antecedent PPT wet-dry tertiles* (Model 3: Panel D) for North Carolina (2008-2015) controlling for same-day ambient mean temperature, public holidays, day-of-week (DOW), county, controls for long-term trends and seasonality. In all panels, the results not controlling for season are shown on the left (red) and results of the sensitivity analysis controlling for meteorological season (spring, summer, fall (reference), winter) are shown on the right (blue). Overall, precipitation was associated with a decrease in all-cause AGI ED rates in North Carolina for absolute precipitation, extreme precipitation, and antecedent wet-dry days (relative to 1 consecutive wet/dry days), with the exception of 8-week wet-dry tertiles of antecedent precipitation, where dry periods were associated with increases in AGI ED rates. Compared to not controlling for season (red), the effect of precipitation when controlling for season (red) yielded slightly larger (up to 1%) decreases in AGI ED rates, but these results were non-significant and small in comparison to the overall effect sizes.

In Model 1, the cumulative association between daily total precipitation (*Absolute PPT*) and AGI rates was estimated with a 0-7 day distributed lag model (DLM) using a 3rd degree polynomial and evaluated at 1, 10, 20, 40, 60, and 80 mm of precipitation (reference: 0 mm). Absolute precipitation was cumulatively associated with 1% decrease in AGI ED rates at 10 mm of daily precipitation over 7 days (0.99; 95% CI: 0.99, 1.00), 3% decrease at 40mm (0.97; 0.95, 0.98), and 6% decrease at 80mm (0.94; 0.90, 0.97) (FIGURE 1, Panel A).

The results of extreme precipitation (Model 2) are displayed in FIGURE 1 (Panel B) as the cumulative association between four measures of 90th, 95th, and 99th percentile (right to left) extreme precipitation (*Extreme PPT*) over 7 days (DLM using 3rd degree polynomial) and AGI ED rates, with all-day (*PPT* ≥ 0mm) and wet-day (*PPT* >0mm) precipitation indices shown respectively on the left and right and the spatial reference of statewide and ZIP code-specific percentiles shown respectively on the top and bottom panels. Over an 8-day distributed linear lag, extreme precipitation decreased AGI ED rates by 1 to 18% depending on the extreme precipitation measure, with the exception of non-seasonal 90th percentile all-day statewide precipitation using an 11 mm cutpoint (1.00; 0.98, 1.01), which had a non-statistically significant null effect. As the extreme precipitation percentile increased from 90th to 99th percentile (left to right), the magnitude of the inverse cumulative association with AGI ED rates also increased for both all days (all) and wet days (wet) and in order of the precipitation cutpoints (mm) represented by the percentiles. For example, results (RR; 95% CI) of the statewide percentile (pct) indices by precipitation cutpoint size (Supplementary FIGURE 2) without controlling for season (top panel, red) were as follows: all-day 90th pct at 11mm (1.00; 0.98, 1.01), all-day 95th pct at 20 mm (0.94; 0.92, 0.95); wet-day 90th pct at 26 mm (0.92; 0.91, 0.94), wet-day 95th pct at 36 mm (0.90; 0.88, 0.93), all-day 99th pct at 43 mm (0.87; 0.83, 0.91), and wet-day 99th pct at 63 mm (0.85; 0.8, 0.91). Using ZIP code-specific extreme precipitation percentiles (bottom panel) led to an additional decrease of AGI ED rates by 1%-3% of extreme precipitation over 7 days when compared to statewide percentiles (top panel) for the index type (all-day or wet-day) and same percentile (90th, 95th, 99th).

To investigate the effect of antecedent precipitation (Model 3), we compared to measures of *Antecedent PPT* modified for daily time series: cumulative wet-dry days and 8-week wet-dry tertiles

(FIGURE 1, Panels C and D respectively). FIGURE 1 (Panel C) illustrates that all categories antecedent precipitation for both consecutive wet days and dry days (2-3, 4-6, and 7+ days) were associated with large (30-36%) and statistically significant decreases in AGI ED rates relative to 1 wet/dry day, with the exception of the non-significant results for 7+ wet days (0.70; 0.43, 1.14). However, the associations for antecedent consecutive wet and dry days had large confidence intervals, were similar for both wet and dry day categories, and were based on small sample sizes (26 to 424 ZIP code-days). For example, the rate ratios between 2-3 days of consecutive dry days and wet days, respectively, and AGI ED rates were 0.66 (0.58, 0.75) and 0.64 (0.51, 0.81) respectively. We also conducted a sensitivity analysis for antecedent consecutive wet-dry days defined as an 8-day (lags = 0-7 days) *linear* distributed lag model (DLM) of continuous wet-dry days, where positive (+) values were assigned to wet days and negative (-) values to dry days. Unlike the large inverse association for categorical antecedent dry days and wet days, the cumulative association of continuous wet-dry days over 7 days and AGI ED rates showed a small, non-significant linear trend from an inverse relationship for consecutive dry (-) days (14 dry days: 0.99; 0.94, 1.04) to a positive relationship for consecutive wet (+) days (10 wet days: 1.01; 0.98, 1.05) (FIGURE 4). FIGURE 1 (Panel D) displays the results for antecedent precipitation defined as wet-dry tertiles of 8-week cumulative daily precipitation (dry, moderate (reference), and wet), which was associated with a statistically significant 2% increase in AGI ED rates for 8-week dry periods and a statistically non-significant 1% decrease in AGI ED rates for 8-week wet periods relative to periods of moderate rainfall. The positive relationship (2% increase) between antecedent 8-week wet-dry tertiles and AGI ED rates was the only positive association statewide and was of similar magnitude whether the precipitation cutpoints for the 33rd and 67th percentiles were calculated statewide, by county, or ZIP code (see Supplementary TABLE 19): statewide 1.02 (1.02, 1.03); county 1.02 (1.01, 1.02); and ZIP code 1.02 (1.02, 1.02).

To investigate the effect modification of antecedent precipitation on the association between extreme precipitation and AGI ED rates (Model 4) statewide in NC, we modeled the interaction between same-day (lag=0) extreme precipitation—defined as greater than or equal to the 95th percentile of daily total precipitation (4 measures: all days vs. wet days; statewide vs. ZIP code-specific)—and 8-week antecedent precipitation wet-dry tertiles, matching the spatial reference areas (statewide and ZIP code-specific only; excluding county-specific antecedent wet-dry tertiles) used to determine the cutpoints for extreme precipitation percentile and antecedent wet-dry tertiles (e.g., statewide 95th extreme with statewide wet-dry tertiles). TABLE 24 shows the association between AGI ED rates and the interaction of same-day 95th percentile extreme precipitation and wet-dry tertiles of antecedent precipitation, with and without controlling for season. Relative to the reference category of same-day non-extreme (<95th percentile) precipitation days following a moderate period of rain (Not Extreme-Mod), days with 95th extreme precipitation were inversely associated with AGI ED rates (1% to 6% decrease depending on category) regardless of antecedent precipitation periods (dry, moderate, or wet) and controlling for season. The largest decreases in AGI ED rates were observed for the statewide wet-day 95th percentile extreme PPT days (PPT $\geq$ 36 mm) following an 8-week dry period (no season: 0.94; 0.91, 0.98; season: 0.94; 0.91, 0.97). Conversely, non-extreme days with rain less than the 95th percentile were positively associated (2% to 7% increase depending on category) with AGI ED rates following an 8-week dry or wet period. The largest increases in AGI ED rates were similarly observed for the statewide wet-day 95th percentile extreme PPT days (PPT $\geq$ 36 mm) following an 8-week dry period (no season: 1.06; 1.02, 1.10; season: 1.07; 1.03, 1.10). Supplementary TABLE 20 presents an alternative version of extreme and antecedent precipitation interaction results reorganized by antecedent wet-dry tertile categories (wet vs. dry) and 95th percentile index types (all-day vs. wet -day) to facilitate comparisons between the results of the effect modification and extreme precipitation or antecedent precipitation alone, such that:

- TABLE 20 Panel A (Model 2): 0-7 day DLM *Extreme PPT* only [ref. FIGURE 1, Panel B],
- TABLE 20 Panel B (Model 3) : 8-week wet-dry tertiles of *Antecedent PPT* only [ref. FIGURE 1, Panel D], and
- TABLE 20 Panel C (Model 4): the interaction between same day 95th percentile and wet-dry tertiles of antecedent precipitation (*Extreme PPT* x Antecedent PPT) [ref. TABLE 24].

3.5. Discussion

The average annual risk of an ED visit for an AGI episode was roughly 3% for North Carolina residents during the study period. Extreme rainfall events were associated with a slightly reduced risk of AGI statewide. Statewide for NC, we observed cumulative 1% to 18% decreases in all-cause AGI ED rates associated with precipitation (both absolute and extreme precipitation measures) over a 0-7-day lag. The inverse relationship between extreme precipitation and AGI ED rates was consistent across extreme precipitation definitions, though the magnitude of the effect depended on index type (all-day, wet-day), spatial reference (statewide, ZIP code-specific), and percentile (90th, 95th 99th). Similarly, we observed a decrease in AGI ED rates (30% to 36%) when antecedent precipitation was defined by consecutive wet-dry days (Tornevi et al., 2013, 2015) for all lengths of consecutive dry days and wet days greater than one. Conversely, when antecedent precipitation was defined as 8-week wet-dry tertiles (Carlton et al., 2014b; D. Lee et al., 2019), we observed a 2% increase in AGI ED rates after dry periods relative to moderately wet periods across all precipitation definitions (statewide, county-specific, and ZIP code specific), though antecedent wet periods had little to no effect. From the effect measure modification (EMM) of 8-week wet-dry tertile antecedent precipitation on 95th percentile same-day extreme precipitation, we observed that all-cause AGI ED rates increased by 2% to 7% in association with non-extreme precipitation following wet or dry periods, and decreased by 1% to 6% for extreme precipitation following dry, moderate, or wet periods. Overall, positive relationships between all-cause AGI ED rates and precipitation were only observed in association with 8-week dry periods of antecedent precipitation and daily non-extreme precipitation (<95th percentile) following dry or wet periods. Sensitivity analyses controlling for season had a relatively small effect, if any, generally increasing the magnitude of a rate ratio between 0% to 1% across all precipitation measures.

One of the earliest studies on extreme precipitation and waterborne disease found that extreme precipitation events above the 90th (80th) percentile preceded 51% (68%) of waterborne disease outbreaks in the United States between 1948 to 1994, controlling for season and hydrologic region (Curriero et al., 2001). Since then, more studies have included extreme precipitation as an exposure measure for AGI and diarrheal illnesses. When considered in the context of the dilution-concentration hypothesis, the consistent statewide inverse association between extreme precipitation and AGI ED rates, with the exception of the neutral and non-significant association for all-day 90th percentile precipitation with a cutpoint of 11 mm), could be the result of a dilution effect on average (Kraay et al., 2020; K. Levy et al., 2016; K. Levy, Hubbard, Nelson, et al., 2009; Moors et al., 2013). Should the trend continue, we hypothesize that lower extreme precipitation thresholds (e.g., 80th percentile) could have positive associations with AGI ED rates. We observed that magnitude of the inverse association for all-day and wet-day 90th, 95th, and 99th extreme precipitation (with and without seasonal controls) increased with respect to increases in the precipitation cutpoints (mm) for the extreme precipitation percentiles, irrespective of the exposure definitions. These estimates were most sensitive to all-day and wet-day exposure definitions and precipitation cutpoints and somewhat sensitive to the spatial reference area (statewide vs. ZIP code) defining the cutpoints such that ZIP code-specific percentile definitions tended to have larger magnitude effect estimates, but were not sensitive to controlling for season. The seasonal sub-analysis could be improved by interacting season with extreme precipitation to test for EMM and testing different definitions of season such as cold and warm (Gleason et al., 2017) and rainy versus dry seasons (Kraay et al., 2020).

In their systematic review of AGI and heavy rainfall, temperature, flooding and drought studies, Levy and colleagues (2016) found that the studies of heavy rainfall and diarrheal disease also had a wide variety of exposure definitions and, of the 10 studies (14 analyses) with quantitative analyses, a significant positive association was reported by 71% (10), a significant negative association by 3 (21%), and no effect by 1 (7%). Similarly, in a systematic review of extreme precipitation or temperature and waterborne infections related to drinking water (DW), Guzman Herrador and colleagues (2015) found 20 studies with positive associations between extreme precipitation and waterborne infections in which 55% (11) of the studies had significant positive associations, 3 (15%) had heterogenous results, and 30% (6) had no association. More recently, Kraay and colleagues (2020) reviewed 19 studies that used extreme precipitation as an exposure measure: 13 of these studies were included in the qualitative synthesis and quantitative meta-analysis (Bhavnani et al., 2014; K. Bush et al., 2014a; K. F. Bush et al., 2014b; Carlton et al., 2014; Chhetri et al., 2017; Gleason & Fagliano, 2017; Jagai et al., 2015; Kang et al., 2015; Mertens et al., 2019; Mukhopadhyay et al., 2019; Phung, Huang, et al., 2015; P. Wang et al., 2018; J. Wu et al., 2014); and 6 were only included in the qualitative synthesis (Bradatan et al., 2020; Cambrea et al., 2019; Mukabutera et al., 2016; Phung et al., 2017; Poulsen et al., 2018). Kraay and colleagues (2020) found the following proportions (counts/739) for inverse, neutral and positive associations amongst the 739 individual analyses for extreme rain from 16 studies: 3.2% (24) inverse associations, 69.3% (512) neutral associations, and 27.5% (203) positive associations. Overall, the authors did not find a statistically significant association between extreme rain and diarrhea from their pooled estimates without modification for antecedent precipitation (IRR: 1.16; 95% CI: 0.946, 1.42). However, when stratified by threshold type, there were non-significant differences in the direction of the pooled estimates (IRR; 95% CI) by percentile threshold, such that the 80th percentile (1.36; 0.884, 209) trended towards a positive association and the 90th (0.978; 0.887, 1.08) and 95th (0.872; 0.877, 1.08) percentiles trended towards negative associations (the 99th percentile pooled estimate had a neutral association (1.00; 0.895, 1.12). The trends of the pooled estimates stratified by percentile type (Kraay et al., 2020) and the trend of the extreme precipitation associations in this study suggest that including percentiles lower than the 90th would be worthwhile in a sensitivity analysis.

When extreme precipitation was modified for antecedent precipitation (K. F. Bush et al., 2014b; Carlton et al., 2014; Chhetri et al., 2017; Mertens et al., 2019), Kraay and colleagues (2020) observed the concentration dilution hypothesis was only partially supported; diarrheal risk statistically significantly increased after dry periods based on pooled estimates (1.26; 1.05, 1.51) (contrary to our results), but the inverse association between extreme rain following wet periods, which we observed, was not statistically significant in the pooled estimates (0.911; 0.771, 1.08). Including the four aforementioned studies reviewed by Kraay and colleagues (2020), there are eight recent climate-diarrhea studies that have tested for effect measure modification (EMM) of extreme precipitation by various measures of antecedent precipitation (K. F. Bush et al., 2014b; Carlton et al., 2014; Chhetri et al., 2017; Graydon et al., 2022; D. Lee et al., 2019; Mertens et al., 2019; Tornevi et al., 2013, 2015). We observed that 95th percentile all-day extreme precipitation was inversely association with AGI ED rates regardless of antecedent period (wet, moderate, or dry) and positively associated with AGI ED rates for wet and dry periods, relative to non-extreme events following moderate periods. Our findings are compared to the statistically significant (*) and non-statistically significant positive and inverse associations found in six of the studies that defined antecedent wet and/or dry periods (excluding Tornevi et al., 2013, 2015) in TABLE 13. With the exception of our study, most studies found at least one statistically significant positive association for extreme precipitation following dry periods, consistent with the concentration effect. However, in the cases of extreme precipitation following wet or moderate periods the results are less consistent and at least one study in each category observed statistically significant positive associations and inverse associations. While there are relatively few studies who have developed (varied) antecedent precipitation measures and tested for EMM of extreme precipitation by antecedent precipitation and there could be many sources of intra-study variation, further development, testing, and standardization of antecedent precipitation measures and EMM with extreme precipitation is an interesting and important area of continued research.

We compared two different antecedent precipitation measures defined as categorical indicator variables at the daily level to capture prior rainfall levels: 8-week wet-dry tertiles (Carlton et al., 2014; D. Lee et al., 2019) and consecutive wet-dry days (Tornevi et al., 2013, 2015). Consecutive wet dry days defined based on a > 0 mm threshold for rainy days (categories: 1 wet/dry, 2-3 dry, 2-3 wet, 4-6 dry, 4-6 wet, 7+ dry, 7+wet) was a poor measure for North Carolina because the majority of observations (over 2 million) were categorized into 1 consecutive wet or dry day, unlike the consecutive wet-dry day categories developed by Tornevi and colleagues (2013, 2015) for Gothenburg, Sweden (1-2 wet/dry, 3-7 dry, 3-7 wet, 8-13 dry, >7 wet, >13 dry). The wet-dry day exposure measure may be less appropriate for locations that rain more frequently, resulting in shorter periods of consecutive dry or wet day. Furthermore, defining wet day based on the threshold > 0mm precipitation (44% wet days), instead of a higher threshold such as > 0.1 mm (39% wet days) or > 1mm, may have misclassified dry days as wet days and interrupted the length of consecutive wet or dry days more frequently. The consecutive wet-dry days exposure variable covered a variable length of time for prior rainfall levels, was more complex to categorize and customize for a given location, and it is unclear whether consecutive days or the cumulative effect of wet or dry days over a period of time is more important. By contrast, the 8-week wet dry tertiles performed well in antecedent precipitation only (Model 3) and EMM of extreme precipitation by antecedent precipitation (Model 4), was simpler to calculate, and covers a fixed length of time. The antecedent precipitation measure developed by Chhetri and colleagues (2017) and adapted for different Great Lakes city locations by Greydon and colleagues (2022) offers a promising alternative. They classify dry and wet periods by normalizing the threshold for a number of days in a 60-day prior period to a given location in order to achieve an even distribution of weeks classified as wet and dry periods. Alternatively, a frequency-based precipitation index over a period of time, similar to degree days, could be used.

TABLE 13. Categorization of association (positive, inverse) for studies testing for effect measure modification extreme precipitation (extreme, not extreme) by antecedent precipitation (wet, moderate, dry periods) on the association with AGI (K. F. Bush et al., 2014; Carlton et al., 2014; Chhetri et al., 2017; Graydon et al., 2022; D. Lee et al., 2019; Mertens et al., 2019) and this study (D).

Extreme	**Antecedent**	**Positive Association***	**Inverse Association***
Yes	Wet	$L^{*}_{a,c}$, $M^{*}_{3\text{-week}}$	B, $Ca^{*}_{2\text{-week}}$, D*, $M_{1\text{-}2\text{-week}}$
Yes	Moderate	$L^{*}_{a,c}$, $M_{1\text{-}2\text{-week}}$	B, $Ca_{2\text{-week}}$, D*
Yes	Dry	B*, Ch, $Ca^{*}_{2\text{-week}}$, $G^{*}_{Ham,3\text{-}5\text{-week}}$, $G^{*}_{Tor,1\text{-week}}$ $L^{*}_{a,c;\ env.\ only}$, $M^{*}_{2\text{-}3\text{-week}}$	D*, $G_{Mil,0\text{-}1\text{-weeks}}$
No	Wet	D*	L_{n}
No	Moderate	reference	reference
No	Dry	D*	$L^{*}_{a,c}$

Notes: No association (not shown)
* Statistically-significant association
Study abbreviations by abbreviated first author last name: B (Bush et al., 2014b), Ca (Carlton et al., 2014), Ch (Chhetri et al., 2017), D (Downs, this study), G (Graydon et al., 2022), L (Lee et al., 2019), M (Mertens et al., 2019)
Strata abbreviations: a (all counties), c (coastal counties), n (northern counties), env. only (environmental serovars), Ham (Hamilton), Mil (Milwaukee), Tor (Toronto Island), x-week (x week lag)

The results from the extreme precipitation (Model 2), antecedent precipitation (Model 3), and EMM of antecedent precipitation on extreme precipitation (Model 3) are more compelling and informative than the association between absolute precipitation and AGI ED rates. Kraay and colleagues (2020) reviewed 50 rainfall studies, including 15 of those studies in their quantitative meta-analysis. The authors found no linear association between rain and diarrhea in the pooled estimates (0.998; 0.967, 1.03) of linear measures of rain (a.k.a. absolute precipitation). Out of 48 studies using rain as the exposure measure, 58% (28) had positive associations with diarrhea. Of the 333 analyses from these studies, 11.7% (39) had inverse associations with rain, 79.7% (599/752) had neutral associations, and 16.1% (121) had positive associations. We observed a small inverse cumulative association between daily total precipitation over an 8-day lag, which is consistent with the minority of rainfall analyses (39). Though less common in existing studies, exploring nonlinear relationships with absolute rainfall measures may better capture the dynamics between absolute precipitation and AGI. Of the reviewed studies, 5 studies had varying types of nonlinear associations with diarrheal risk: increases at high and low rainfall levels (U-shaped) (Fang et al. 2019; Dunn and Johnson 2018; Ikeda et al. 2019), moderate rainfall (Chowdhury et al. 2018), and high levels only (Uejio et al. 2014). By using a (linear) distributed lag model (DLM) in this study, we were unable to test whether there was a non-linear relationship between absolute rainfall and AGI ED rates, a likely possibility because extreme precipitation had a larger association with AGI ED rates. A distributed lag nonlinear model (DLNM) may better specify this association (Gasparrini, 2011; Gasparrini et al., 2010).

To the best of our knowledge, this is the first North Carolina study to investigate the relationship between AGI and weather, particularly precipitation, using time series regression and the first systematic exploration of the sensitivity of the association between of precipitation and AGI to different precipitation measures. We know of only one other investigation on rainfall and AGI in North Carolina, the Master's thesis by Hartley (2016), which used bivariate and spatial analyses. Using all-cause AGI ED visits with no nausea or vomiting derived from NC DETECT (2008-2012), Hartley found that the average number of ED visits per day per 100,000 person-years per county was higher after periods of heavy precipitation (greater than 2 inches, equivalent to 50.8 mm) compared to light precipitation (less than 2 inches) for cumulative rainfall calculated over 0-3 day and 0-10-day lags. Hartley also calculated and mapped the average number of ED visits per day per 100,000 person-years for by county for days by 3-day and 10-day lag heavy and light precipitation, and the proportional difference between the ED rates for heavy and light precipitation. Hartley found statistically significant spatial clustering within North Carolina, as indicated by *Moran's I* spatial autocorrelation, of the proportional differences between light and heavy precipitation, suggesting regional differences in the relationship between AGI and precipitation. To extend and improve upon Hartley (2016), we employed more advanced statistical techniques at a higher spatial resolution (ZIP code level) over a longer study period (2008-2015), explored a range of precipitation cutpoints, and were able to control for short-term and long-term time trends, seasonality, and county through time series regression and distributed lag models (DLM). Unlike Hartley (2016), we observed an overall inverse association between extreme precipitation and AGI ED rates statewide for North Carolina.

Also relevant to our study are three recent statewide climate-diarrhea environmental epidemiology studies in the southeast U.S. on AGI and hurricane flooding in North Carolina (Quist, Fliss, et al., 2022a), AGI and industrial hog operations in NC (Quist, Holcomb, et al., 2022b), and salmonellosis and extreme and antecedent precipitation in Georgia (GA) (D. Lee et al., 2019). Two additional U.S. statewide studies on AGI and extreme precipitation were conducted in the northeast U.S. on AGI ED rates and extreme precipitation in Massachusetts (MA), stratified by type of combined sewer system (Jagai et al., 2015c), and on AGI in-patient hospitalizations and heavy precipitation in New Jersey (NJ), stratified by season and drinking water source (Gleason & Fagliano, 2017d). These studies found the associations between extreme precipitation and all-cause or pathogen-specific AGI sensitive to physiographic region (D. Lee et al., 2019); surface water, groundwater, or other drinking water source (Gleason & Fagliano, 2017d); cold and warm seasons (Gleason & Fagliano, 2017d), sewer system type and characteristics (Jagai et al., 2015c), and suggest that there are likely regional and population-specific differences in the relationship between precipitation and AGI. The need for a regional model is further supported by recent findings that areas of North Carolina with high concentrations of industrial hog operation (clustered in southeastern NC) were associated with increases in AGI ED rates following extreme precipitation (Quist, Holcomb, et al., 2022b).

This study's strengths include developing a high-resolution (daily, ZIP code-level) statewide model for North Carolina using time series analysis techniques (quasi-Poisson GLM and DLM) to estimate the statewide association between all-cause AGI ED rates and precipitation in North

Carolina, using multiple precipitation measures (absolute precipitation, extreme precipitation, antecedent precipitation) and definitions. By broadly defining AGI as all-cause AGI (including nausea and vomiting) we were able to have sufficient sample size and power over time allowing to use daily time series at higher spatial resolution (ZIP code) and to capture cases of AGI that may have been missed by the limited pathogen-specific and laboratory-confirmed cases. However, by grouping multiple etiologies into all-cause AGI, we are not able to distinguish unique pathogen-specific relationships between AGI and weather (particularly precipitation, temperature, and season). AGI is self-limited and often underreported (Craun et al., 2010; Roy et al., 2006; Scallan, Griffin, et al., 2011; Scallan, Hoekstra, et al., 2011; Schuster et al., 2005; Yoder et al., 2008) and total cases of AGI are only fractionally captured by ED visits and may underestimate and/or not be representative of AGI (Mead et al., 1999). The ED visits in the study sample also excluded patients with using post office boxes for their billing addresses. Although we explored multiple exposure definitions for extreme precipitation and antecedent precipitation, our sensitivity analyses could be improved by accounting for nonlinearity (i.e., compare the DLM to a DLNM) in the relationship between absolute precipitation; assessing additional lag terms for absolute, extreme, and antecedent precipitation, particularly in the interaction between extreme precipitation and absolute precipitation; expanding the sensitivity analysis of extreme precipitation to thresholds below 90th percentile (e.g., 50th, 70th, 80th); and considering an alternate classification of rainy days as > 1 mm in comparison to > 0 mm precipitation. Finally, the estimates of the association between all-cause AGI ED rates and daily precipitation are defined for the entire state of North Carolina and do not include other strata to account for differences in this relationship by sub-population. Geographic and population differences may arise from multiple factors, including region, water source type, socioeconomic status, age, agricultural and livestock exposures.

3.6. Conclusions

We demonstrated that a NC statewide protective effect (inverse association) between 90th, 95th, and 99th percentile extreme precipitation with cutpoints above 11 mm and AGI ED rates was robust across multiple exposure definitions (all-day vs. wet-day; statewide vs. county-specific) with and without controlling for season and for 95th percentile extreme precipitation events following dry, moderate, and wet periods. However, levels of 8-week antecedent precipitation increased AGI risk following dry periods and when same-day non-extreme (<95th percentile) events were preceded by both wet and dry periods. While our results are suggestive of a dilution effect for higher levels of extreme precipitation on average across North Carolina, a statewide model cannot be used to identify at-risk populations. To better develop targeted public health policies and interventions and better understand contextual concentration-dilution effects for North Carolina, a model that accounts for variation in population, regional, and/or environmental characteristics and includes lower levels of heavy precipitation is needed. Our findings offer further support explore the concentration-dilution hypothesis further by developing and testing multiple measures of antecedent precipitation for different levels of spatiotemporal aggregation and at the effect measure modification of and antecedent precipitation. Ultimately, increasing consistency in antecedent precipitation exposure definitions to increase comparability between studies and our understanding of concentration-dilution dynamics across a variety of settings.

We echo the recommendations from Kraay and colleagues (2020) for studies to consider effect modification by antecedent precipitation and report the both precipitation percentiles and the corresponding specification of the numerical precipitation cutpoints. We expand this recommendation and suggest that studies considering extreme precipitation exposures include details about the exposure definition (especially thresholds for and the inclusion or exclusion of zero-precipitation days); consider sensitivity analyses that may include both all-day, wet-day percentiles or, more conservatively, include a range of all-day percentiles up to and beyond the 99th percentile; and explore frequency-based precipitation indices for higher levels of spatiotemporal aggregation that can take advantage of more resolved weather and data. It would be very informative and improve study comparability for researchers to additionally provide the equivalent all-day *and* wet-day percentiles for each precipitation cutpoint. Future research on extreme precipitation and antecedent precipitation measures should be developed and compared across a range of spatiotemporal aggregation levels, particularly for daily, weekly, and monthly time series and individual locations compared to larger spatial extents that may require spatial aggregation.

3.7. Tables

TABLE 14. Frequency of acute gastrointestinal illness (AGI) emergency department (ED) visits in North Carolina by AGI case definition (based on the following ICD-9-CM diagnostic codes: pathogen-specific infectious GI illness 001-009, non-infectious GI illness 558.9, not otherwise specified diarrhea 787.91, nausea and/or vomiting 787.01-787.03), year, patient characteristics (age group, sex, health insurance type), and physiographic region (Mountains, Piedmont, Coast). All-cause AGI (including symptoms of nausea and vomiting) was selected as the outcome because of limited laboratory-confirmed etiologies. All-cause AGI ED visits were aggregated into daily counts of AGI by ZIP code across the state of North Carolina based on ZIP code spatial areas for the study period of January 1, 2008 to September 31, 2014, for which ICD-9 codes were available before the NC healthcare systems switched to ICD-10 codes on October 1, 2015.

	Frequency of Emergency Department (ED) Visits by Acute Gastrointestinal Illness (AGI) Case Definition								
Characteristic	**All-cause AGI,** N = 2,776,870[1]	**All-cause AGI (no NV[2]),** N = 1,079,955[1]	**Bacterial AGI,** N = 46,705[1]	***E. coli* (all),** N = 279[1]	***C. difficile*,** N = 37,176[1]	**Cholera (all),** N = 120[1]	**Viral AGI,** N = 79,379[1]	**Protozoan AGI,** N = 1,157[1]	**Protozoan AGI (no abscesses),** N = 1,023[1]
Year									
2008	304,567 (11%)	119,189 (11%)	4,281 (9.2%)	25 (9.0%)	3,233 (8.7%)	0 (0%)	9,668 (12%)	74 (6.4%)	68 (6.6%)
2009	340,739 (12%)	130,697 (12%)	4,464 (9.6%)	32 (11%)	3,419 (9.2%)	5 (4.2%)	10,618 (13%)	69 (6.0%)	68 (6.6%)
2010	354,769 (13%)	135,926 (13%)	4,995 (11%)	25 (9.0%)	3,790 (10%)	3 (2.5%)	10,565 (13%)	47 (4.1%)	45 (4.4%)
2011	348,020 (13%)	127,555 (12%)	6,005 (13%)	37 (13%)	4,723 (13%)	5 (4.2%)	8,838 (11%)	71 (6.1%)	70 (6.8%)
2012	378,207 (14%)	140,612 (13%)	6,607 (14%)	46 (16%)	5,337 (14%)	4 (3.3%)	10,572 (13%)	65 (5.6%)	63 (6.2%)
2013	391,633 (14%)	156,063 (14%)	7,529 (16%)	56 (20%)	6,079 (16%)	79 (66%)	12,193 (15%)	616 (53%)	519 (51%)
2014	377,830 (14%)	149,378 (14%)	7,056 (15%)	32 (11%)	5,864 (16%)	21 (18%)	8,953 (11%)	169 (15%)	146 (14%)
2015 (Jan-Sept)	281,105 (10%)	120,535 (11%)	5,768 (12%)	26 (9.3%)	4,731 (13%)	3 (2.5%)	7,972 (10%)	46 (4.0%)	44 (4.3%)
(Missing)	0	0	0	0	0	0	0	0	0
Season									
Fall	608,277 (22%)	217,897 (20%)	10,565 (23%)	65 (23%)	8,348 (22%)	28 (23%)	13,698 (17%)	354 (31%)	323 (32%)
Winter	753,757 (27%)	313,824 (29%)	11,510 (25%)	56 (20%)	9,306 (25%)	23 (19%)	26,980 (34%)	204 (18%)	184 (18%)
Spring	756,397 (27%)	313,938 (29%)	12,311 (26%)	71 (25%)	10,017 (27%)	28 (23%)	25,642 (32%)	311 (27%)	272 (27%)
Summer	658,439 (24%)	234,296 (22%)	12,319 (26%)	87 (31%)	9,505 (26%)	41 (34%)	13,059 (16%)	288 (25%)	244 (24%)
(Missing)	0	0	0	0	0	0	0	0	0
Age Group									
0-4	320,687 (12%)	152,616 (14%)	994 (2.1%)	17 (6.1%)	409 (1.1%)	9 (7.5%)	17,772 (22%)	26 (2.2%)	18 (1.8%)
5-11	160,459 (5.8%)	61,984 (5.7%)	596 (1.3%)	9 (3.2%)	174 (0.5%)	14 (12%)	7,921 (10.0%)	170 (15%)	137 (13%)
12-17	115,628 (4.2%)	39,638 (3.7%)	574 (1.2%)	6 (2.2%)	174 (0.5%)	15 (12%)	3,803 (4.8%)	273 (24%)	257 (25%)
18-64	1,799,891 (65%)	650,441 (60%)	20,147 (43%)	123 (44%)	14,130 (38%)	68 (57%)	38,279 (48%)	590 (51%)	516 (50%)
65+	380,159 (14%)	175,258 (16%)	24,393 (52%)	124 (44%)	22,288 (60%)	14 (12%)	11,602 (15%)	98 (8.5%)	95 (9.3%)
(Missing)	46	18	1	0	1	0	2	0	0
Sex									

Characteristic	Frequency of Emergency Department (ED) Visits by Acute Gastrointestinal Illness (AGI) Case Definition								
	All-cause AGI, N = 2,776,870[1]	**All-cause AGI (no NV[2]),** N = 1,079,955[1]	**Bacterial AGI,** N = 46,705[1]	***E. coli* (all),** N = 279[1]	***C. difficile*,** N = 37,176[1]	**Cholera (all),** N = 120[1]	**Viral AGI,** N = 79,379[1]	**Protozoan AGI,** N = 1,157[1]	**Protozoan AGI (no abscesses),** N = 1,023[1]
F	1,783,535 (64%)	664,537 (62%)	28,409 (61%)	178 (64%)	23,146 (62%)	62 (52%)	46,301 (58%)	370 (32%)	328 (32%)
M	993,197 (36%)	415,355 (38%)	18,294 (39%)	101 (36%)	14,028 (38%)	58 (48%)	33,076 (42%)	787 (68%)	695 (68%)
U	65 (<0.1%)	29 (<0.1%)	2 (<0.1%)	0 (0%)	2 (<0.1%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
(Missing)	73	34	0	0	0	0	2	0	0
Health Insurance									
Public	1,327,606 (49%)	535,865 (51%)	30,373 (67%)	163 (60%)	26,088 (73%)	57 (48%)	42,547 (55%)	474 (42%)	423 (42%)
Private	650,939 (24%)	253,852 (24%)	8,749 (19%)	72 (26%)	5,910 (17%)	36 (30%)	18,839 (24%)	426 (37%)	368 (37%)
Self-pay/ Uninsured	569,046 (21%)	202,102 (19%)	3,115 (6.9%)	26 (9.6%)	1,477 (4.1%)	23 (19%)	12,939 (17%)	193 (17%)	175 (17%)
Other/ Unknown	160,189 (5.9%)	61,494 (5.8%)	2,801 (6.2%)	11 (4.0%)	2,288 (6.4%)	4 (3.3%)	3,149 (4.1%)	47 (4.1%)	40 (4.0%)
(Missing)	69,090	26,642	1,667	7	1,413	0	1,905	17	17
Physiographic Regions[3]									
Mountains	183,884 (6.6%)	68,581 (6.4%)	3,349 (7.2%)	16 (5.7%)	2,643 (7.1%)	87 (72%)	3,715 (4.7%)	628 (54%)	520 (51%)
Piedmont	1,724,171 (62%)	671,700 (62%)	30,001 (64%)	132 (47%)	24,353 (66%)	22 (18%)	48,964 (62%)	425 (37%)	404 (39%)
Coast	868,815 (31%)	339,674 (31%)	13,355 (29%)	131 (47%)	10,180 (27%)	11 (9.2%)	26,700 (34%)	104 (9.0%)	99 (9.7%)

[1]Frequency

[2]No NV – all-cause AGI excluding nausea (N) and/or vomiting (V) corresponding to ICD-9-CM diagnostic codes 787.01-787.91.

[3]Physiographic regions were defined by the maximum spatial overlap between ZIP codes and a high resolution physiographic map from the North Carolina Department of Environmental Quality (NCDEQ) (https://data-ncdenr.opendata.arcgis.com/maps/ncdenr::physiographic-provinces-of-nc/explore). North Carolina physiography can be categorized in multiple ways: shown are 3-category (Mountains, Piedmont, Coast) and 4-category (Mountains or Blue Ridge, Piedmont, Inner Coastal Plain, and Outer Coastal Plain) regions.

TABLE 15. Descriptive statistics for main outcome (all-cause AGI ED visits) and meteorological exposures (temperature, T, and precipitation, PPT) at the ZIP code-day level for 2008-September 2015. Precipitation measures include absolute precipitation in mm (PPTmean, PPTmin, PPTmedian, PPTmax), statewide and ZIP code-specific all-day and wet-day percentiles of 90th, 95th, and 99th percentile Extreme PPT, antecedent precipitation indicator variables for consecutive wet (+) and dry (-) days following Tornevi (2013, 2015) and for 8-week cumulative precipitation tertiles (Dry, Moderate (Mod), Wet) by spatial unit (state, county, ZIP code) following Carlton et al. (2014) (statewide) and Lee et al. (2019) (county- and ZIP code-specific) (2008-2015). A given weather variable was calculated from the spatial mean, maximum (max), or minimum (min) across each ZIP code from gridded PRISM data (4x4-km grids downscaled to 1x1-km grids). The time series includes 2830 days and 737 ZIP codes.

Variable (model predictors are bolded)		*Value per ZIP code-day*									
AGI (outcome)	**N > 0 (%)**	**Min**	**25th**	**50th**	**Mean**	**75th**	**Max**	**SD**			
All-cause AGI ED visits	1,052,347 (50.7%)	0.0	0.0	1.0	1.3	2.0	31.0	2.2	-	-	-
TEMPERATURE (T) [1]	**Total (n)**	**Min**	**25th**	**50th**	**Mean**	**75th**	**Max**	**SD**			
Mean Temperature, Tmean (°C)	2,085,710	-18.0	8.1	16.4	15.3	23.2	32.9	8.9	-	-	-
ABSOLUTE PRECIPITATION (PPT)[2]	**Total (n)**	**Min**	**25th**	**50th**	**Mean**	**75th**	**Max**	**SD**			
PPTmean (mm), all days	2,084,236	0.0	0.0	0.0	3.4	2.2	248.5	9.0	-	-	-
PPTmean (mm), on rainy days: PPTmean >0mm (excluding dry days, 56%)	929,416	0.0	0.6	3.1	7.7	9.9	248.5	12.2	-	-	-
PPTmean (mm), on rainy days: PPTmean >0.1mm (excluding dry days, 61%)	823,214	0.1	1.2	4.2	8.7	11.3	248.5	12.6	-	-	-
EXTREME PRECIPITATION (PPT)	**Total (n)**	**≥ 90th**			**≥ 95th**			**≥ 99th**			
		n (%)	**Mean**	**SD**	**n (%)**	**Mean**	**SD**	**n (%)**	**Mean**	**SD**	
All-day (PPTmean ≥ 0 mm)											
Statewide cutpoint	2,084,236	210,462 (10.1%)	285.6	27.1	102,614 (4.9%)	139.2	18.6	19,816 (1.0%)	26.9	9.8	-
ZIP code-specific cutpoints	2,084,236	210,614 (10.1%)	285.8	4.9	102,927 (4.9%)	139.7	4.3	19,882 (1.0%)	27.0	2.1	-
Wet-day (PPTmean > 0 mm)											
Statewide cutpoint	2,084,236	88,490 (4.2%)	120.1	17.8	43,228 (2.1%)	58.7	13.9	8,725 (0.4%)	11.8	6.1	-
ZIP code-specific cutpoints	2,084,236	88,746 (4.3%)	120.4	15.0	43,468 (2.1%)	59.0	8.2	8,897 (0.4%)	12.1	2.2	-

ANTECEDENT PRECIPITATION (PPT)	Total (n)	Min	25th	50th	Mean	75th	Max	SD	Dry (n)	Mod (n)	Wet (n)
Wet-dry Days (Tornevi, 2013, 2015)[3]											
Antecedent Consecutive Wet (+) & Dry (-) Days	1,815,231	-14	-1	-1	-0.1	1	10	1.0	1,154,820	-	929,416
Consecutive days:											
Wet days (dry days excluded)	1,815,231	1	1	1	1.0	1	10	0.06	-	-	929,416
Dry days (wet days excluded)	1,815,231	1	1	1	1.0	1	14	0.09	1,154,820	-	-
Wet-dry Tertiles (Carlton et al., 2014; Lee et al., 2019)											
Antecedent 8-week cumulative PPT tertiles by spatial unit:											
Statewide cutpoint	2,021,591	-	-	-	-	-	-	-	620,841	735,543	665,207
County-specific cutpoints	2,021,591	-	-	-	-	-	-	-	618,924	735,721	666,946
ZIP code-specific cutpoints	2,021,591	-	-	-	-	-	-	-	617,631	737,156	666,804

(n) – count of ZIP code-days

[1] The spatial mean of mean ambient temperature (*Tmean*) was calculated across ZIP code by day.

[2] PPTmean is the spatial mean of daily total precipitation (PPT) for all 1x1-km grids of PRISM data whose centroids lie within the ZIP code polygon (2013) and corresponds to PPT unless stated otherwise.

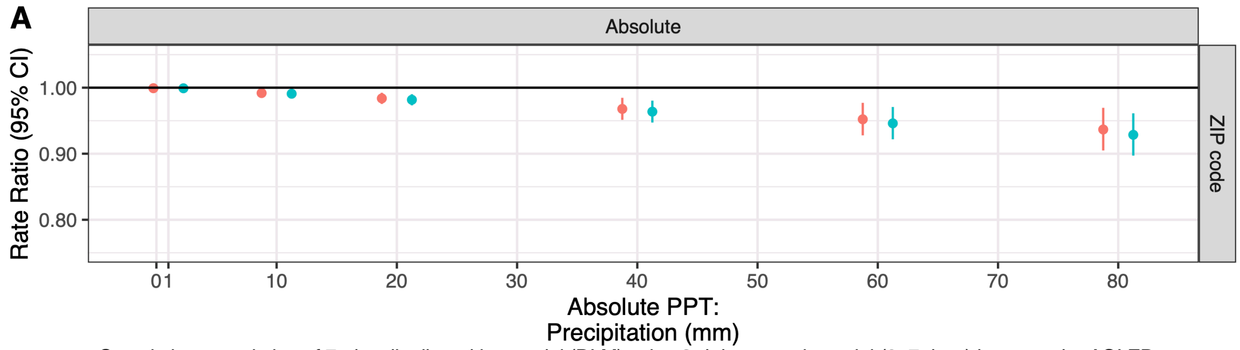
[3] Consecutive wet and dry days were defined as having daily total precipitation greater than 0 mm. Consecutive wet and dry days prior to each day were classified into 7 categories: 1 Wet/Dry (n = 2,084,623), 2-3 Dry (n= 424), 2-3 Wet (n= 349), 4-5 Dry (n= 262), 4-5 Wet (n= 124), 7+ Dry (n= 83), and 7+ Wet (n= 26). Counts of day-ZIP codes corresponding to the total number of consecutive dry and wet days are displayed in Supplementary TABLE 18.

TABLE 16. Association between all-cause AGI ED rates and the combined interaction effect of same-day (lag=0) Extreme (95th percentile) and Antecedent (8-week wet-dry tertiles) precipitation (PPT) across North Carolina from January 1, 2008 to September 30, 2015. Extreme precipitation was defined as an indicator variable for daily total precipitation greater than or equal to the 95th percentile. Cutpoints (mm) defining extreme precipitation varied by all (PPT >=0 mm) vs. wet days (PPT > 0mm) and the spatial area over which the range of precipitation values was defined (statewide vs. ZIP code-specific). The interaction effect was also compared with and without controlling for 4-category meteorological season.

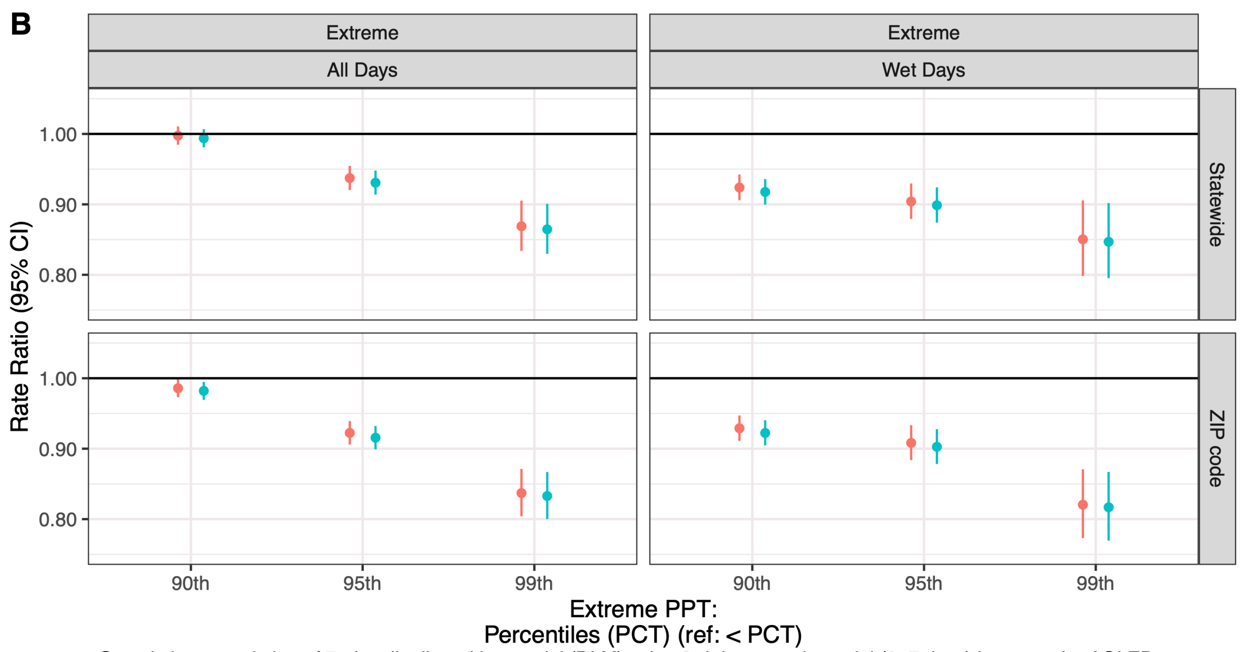
			Precipitation cutpoints defined by all/wet days and statewide/ZIP code			
			All Days (PPT ≥ 0 mm) RR (95% CI)		**Wet Days** (PPT > 0 mm) RR (95% CI)	
Extreme PPT (95th PCT, lag = 0)	**Season (ref: fall)**	**Antecedent PPT (8-week wet-dry tertiles)**	**Statewide**	**ZIP code**	**Statewide**	**ZIP code**
Extreme: 95th	*No Season*	*Dry*	0.97 (0.96, 0.99)	0.98 (0.96, 0.99)	0.94 (0.91, 0.98)	0.96 (0.92, 0.99)
		Mod	0.99 (0.98, 1.00)	0.98 (0.97, 0.99)	0.98 (0.96, 1.00)	0.97 (0.96, 0.99)
		Wet	0.98 (0.97, 0.99)	0.98 (0.98, 0.99)	0.98 (0.97, 0.99)	0.99 (0.97, 1.00)
	Season	*Dry*	0.97 (0.95, 0.99)	0.97 (0.96, 0.99)	0.94 (0.91, 0.97)	0.95 (0.92, 0.98)
		Mod	0.99 (0.98, 1.00)	0.98 (0.97, 0.99)	0.98 (0.96, 1.00)	0.97 (0.95, 0.99)
		Wet	0.98 (0.97, 0.99)	0.98 (0.98, 0.99)	0.98 (0.97, 0.99)	0.99 (0.97, 1.00)
Not Extreme	*No Season*	*Dry*	1.03 (1.01, 1.04)	1.02 (1.01, 1.04)	1.06 (1.02, 1.10)	1.05 (1.01, 1.08)
		Mod	Ref	Ref	Ref	Ref
		Wet	1.02 (1.01, 1.03)	1.02 (1.01, 1.03)	1.02 (1.01, 1.03)	1.02 (1.00, 1.03)
	Season	*Dry*	1.03 (1.01, 1.05)	1.03 (1.01, 1.05)	1.07 (1.03, 1.10)	1.05 (1.02, 1.09)
		Mod	Ref	Ref	Ref	Ref
		Wet	1.02 (1.01, 1.03)	1.02 (1.01, 1.03)	1.02 (1.01, 1.03)	1.02 (1.00, 1.03)

Cumulative RR (95% CI) of the effect of 95th percentile Extreme precipitation (PPT) and wet-dry tertiles of 8-week prior Antecedent PPT (Dry, Moderate (Mod), Wet) on all-cause AGI. The spatial aggregation level for precipitation cutpoints was matched between Extreme PPT and Antecedent PPT measures (i.e., statewide and ZIP code-specific).

3.8. Figures

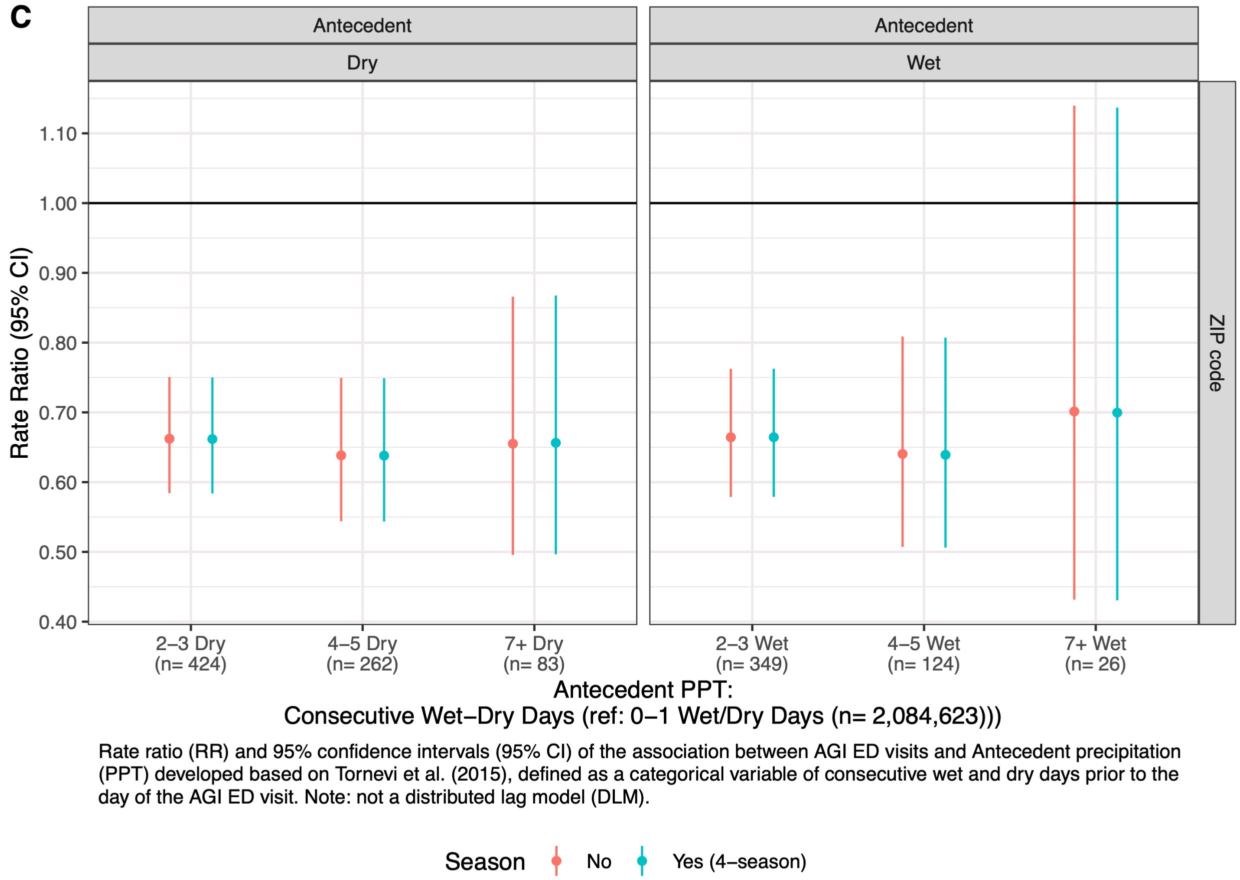


Cumulative association of 7–day distributed lag model (DLM) using 3rd degree polynomial (0–7 days) between the AGI ED visit rate and daily total (Absolute) precipitation (PPT), evaluated at 1, 10, 20, 40, 60, and 80 mm of daily total precipitation (ref = 0 mm).



Cumulative association of 7–day distributed lag model (DLM) using 3rd degree polynomial (0–7 days) between the AGI ED visit rate and an Extreme precipitation (PPT) event, at which daily total precipitation exceeded the Xth percentile (ref = <Xth percentile).

C



Rate ratio (RR) and 95% confidence intervals (95% CI) of the association between AGI ED visits and Antecedent precipitation (PPT) developed based on Tornevi et al. (2015), defined as a categorical variable of consecutive wet and dry days prior to the day of the AGI ED visit. Note: not a distributed lag model (DLM).

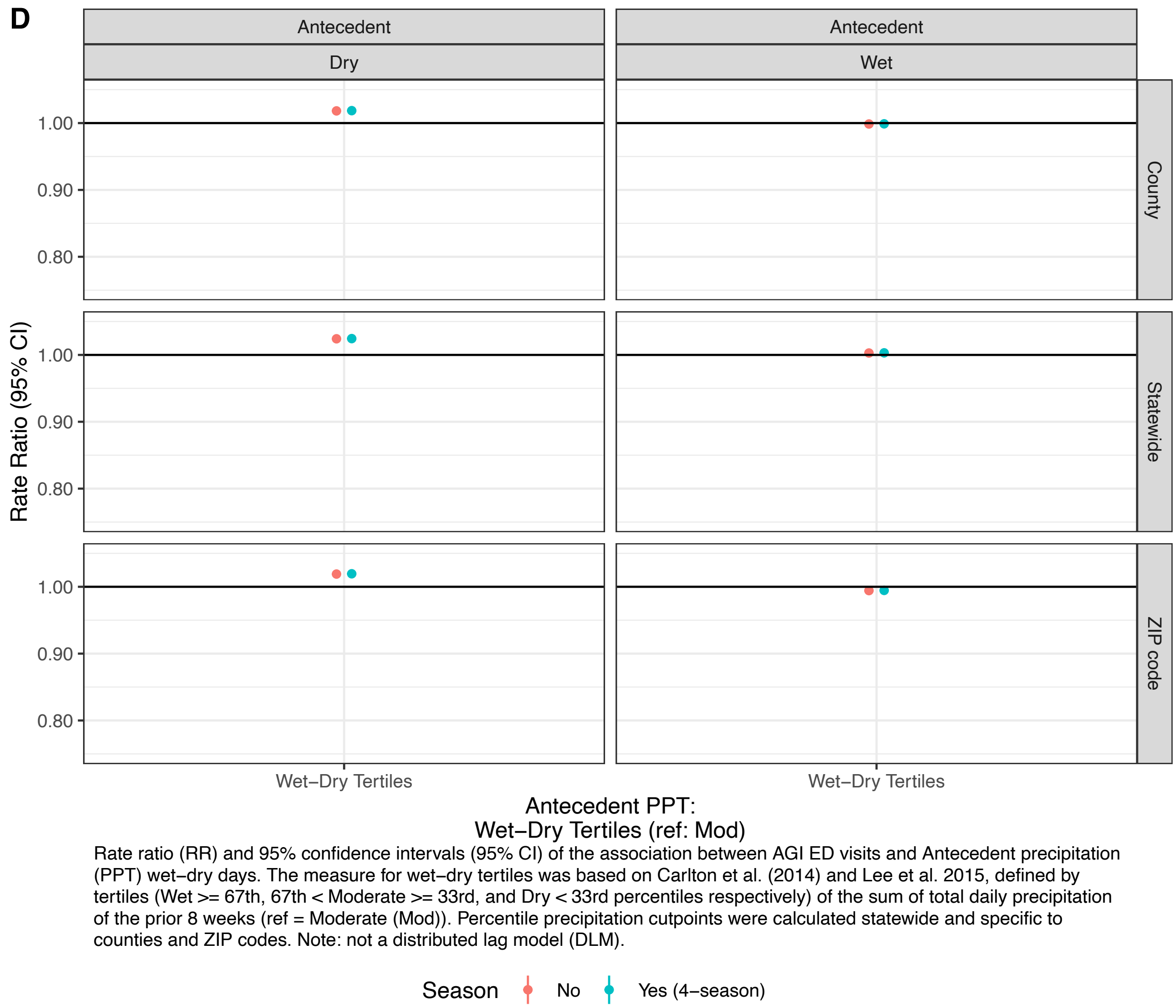


FIGURE 1. The association (rate ratios and 95% confidence intervals) between various precipitation (PPT) metrics and daily all-cause AGI ED rate from time series models at the ZIP code-level across the state of North Carolina (2008-2015). All models have controlled for same-day ambient mean temperature (Tmean), public holidays, day-of-week (DOW), county, controls for long-term and seasonal temporal trends (natural spline of day-of-year (DOY) with 6 degrees of freedom interacted with an indicator for year, ns(DOY, df = 6)/year), and an offset of the log of the annual ZIP code population. Red points indicate the association or cumulative association (for distributed lag models in Panels A and B) without controlling for season. Blue points correspond to results of the seasonal sub-analysis, controlling for the 4 meteorological seasons in the North Hemisphere (ref = fall). Overall, controlling for season slightly increases the strength of the effect for Absolute and Extreme PPT, but does not affect the association for Antecedent PPT.

PANEL A: Absolute PPT: **Continuous measure of the ZIP code spatial mean of daily total precipitation in mm (PPT or PPTmean). The absolute PPT term was modeled as a 7-day distributed lag model (DLM) using a 3rd degree polynomial. The estimated overall cumulative association of PPT on all-cause AGI ED rates over 7 days is evaluated at 1, 10, 20, 40, 60, and 80 mm of daily total precipitation (ref = 0 mm).**

PANEL B: Extreme PPT: **The 90th, 95th, and 99th percentiles of four measures of extreme PPT were developed into a precipitation index using precipitation cutpoints (mm) that varied by all days (≥ 0**

mm) or wet days (> 0 mm) using all ZIP codes statewide versus or only values specific to a given ZIP code, yielding all days-statewide, all days-ZIP code, wet days-statewide, wet days-ZIP code. Extreme PPT was coded into a binary indicator variable based on whether daily total precipitation was equal to or exceed a given cutpoint for the 90th, 95th, or 99th percentile (PCT) of the data subset (*extreme*: $PPT_t \geq$ cutpoint for Xth PCT; *non-extreme,* $PPT_t <$ Xth PCT) the precipitation cutpoints. Extreme PPT was modeled as a 7-day DLM using a 3rd degree polynomial. The rate ratio (95% CI) represents the overall cumulative association of a day classified as having Extreme PPT with all-cause AGI ED rates (ref: PPT_t <Xth PCT).

***PANEL C: Antecedent PPT – Wet-dry Days*: The first measure for Antecedent PPT is wet-dry days, based on Tornevi et al. (2013, 2015). Wet-dry days was defined as a categorical variable of the number of consecutive wet (PPT > 0mm) or dry (PPT = 0 mm) days prior to a given day (t=0) with 7 categories: 1 Wet/Dry (n = 2,084,623), 2-3 Dry (n= 424), 2-3 Wet (n= 349), 4-5 Dry (n= 262), 4-5 Wet (n= 124), 7+ Dry (n= 83), and 7+ Wet (n= 26) consecutive days. Categories were adjusted from Tornevi et al. (2015) to better fit the distribution of consecutive wet-dry days in North Carolina (1-2 wet/dry, 3-7 dry, 3-7 wet, 8-13 dry, >7 wet, >13 dry). The rate ratios (95% CI) represent the association between consecutive wet-dry day categories (ref: 1 day wet/dry) and AGI ED rates using a time series model, not a DLM.**

PANEL D: Antecedent PPT – Wet-dry Tertiles: The second measure for Antecedent PPT is wet-dry tertiles (Dry < 33rd PCT, Moderate, Wet ≥ 67th PCT,) based on the cumulative sum of daily total precipitation over the 8 weeks prior to a given day (t=0), adapted from 8-week tertiles Carlton et al. (2014) and Lee et al. (2019). The rate ratio (95% CI) represents the association between 8-week antecedent Dry or Wet tertiles (ref = Mod) and AGI ED rates for three definitions of wet-dry tertiles based on statewide, county-specific, ZIP code-specific precipitation cutpoints (described in TABLE 17).

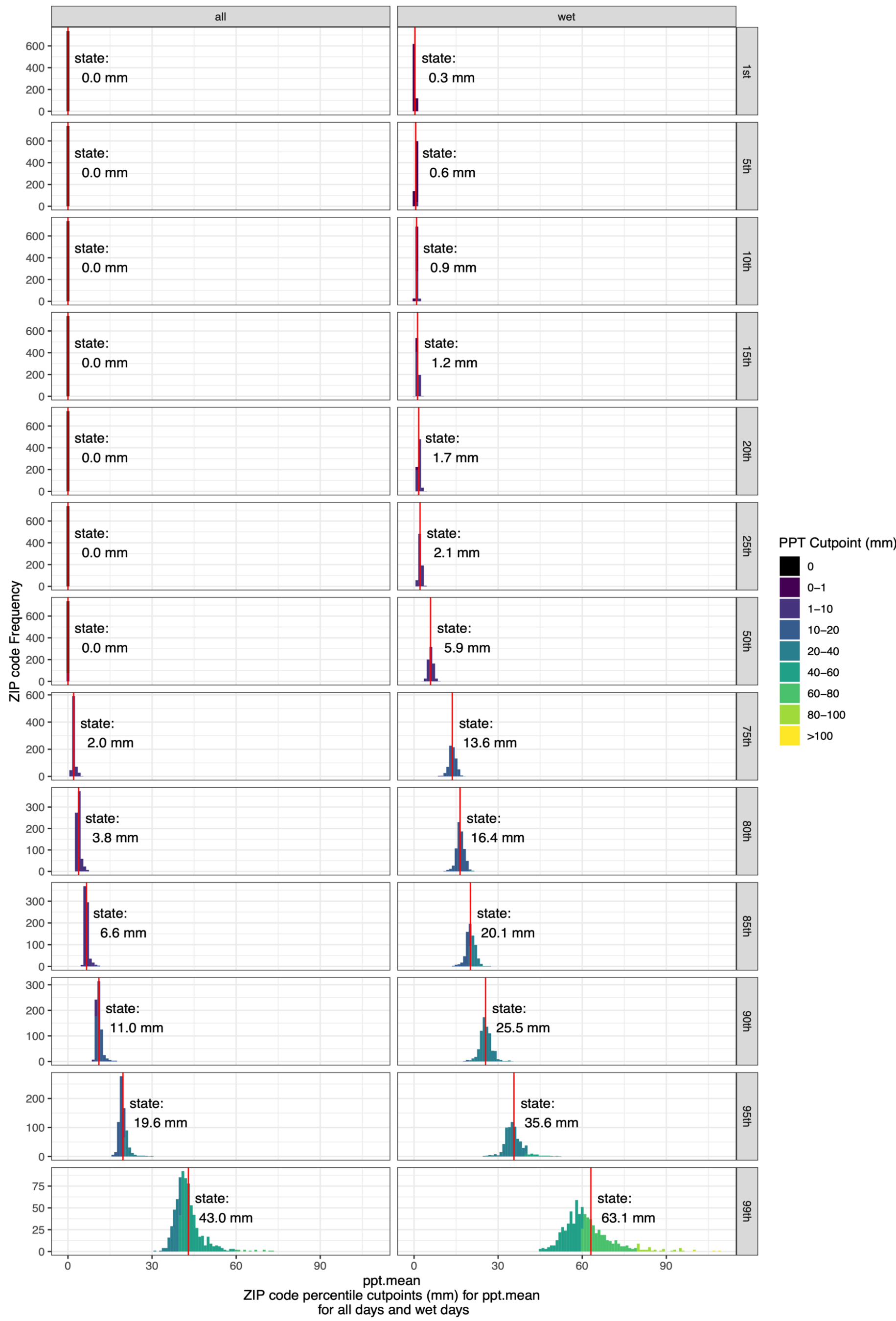


FIGURE 2. Histogram of extreme precipitation (PPT) percentile cutpoints (mm) by NC ZIP code for the spatial mean of daily total precipitation (ppt.mean) as calculated by all-day percentiles (all: ppt.mean ≥ 0 mm) [left panels] and wet-day percentiles (wet: ppt.mean > 0mm) [right panels]. Statewide percentile cutpoints (mm) are indicated by red vertical lines for comparison to ZIP code percentile cutpoints. Extreme precipitation percentiles increase from top to bottom of figure (1st, 5th, 10th, 15th, 20th, 35th, 50th, 75th, 80th, 80th, 90th, 95th, 99th). The 90th, 95th, and 99th percentiles are used to derive daily extreme precipitation indicators used in the time series regression models.

FIGURE 2 shows the distribution (histograms) of the cutpoints (mm) for ZIP code-level daily total precipitation (PPT) for a range of precipitation percentile thresholds (Xth: 1st-99th), increasing from top to bottom. The left column corresponds to all-day percentiles (PPT>=0mm) and the right column to wet-day percentiles (PPT>0mm). Each histogram shows the distribution of frequency of ZIP codes (y-axis) for a given bin of ZIP code-specific precipitation cutpoints (mm) that correspond to the Xth all-day or wet-day precipitation percentile threshold, as well as the NC statewide precipitation cutpoint (mm) (red vertical lines). Four patterns are apparent in FIGURE 2. First, the mean and standard deviation of the histograms increase as the percentile threshold increases. Second, the statewide percentile cutpoints are usually greater than or equal to the mean of the ZIP code-specific cutpoints for both all-day and wet-day percentiles. Third, the statewide and ZIP code-specific cutpoints for all-day percentiles, which include non-rain days (PPT=0mm), are equal to 0 mm beyond the 50th percentile, while the wet-day (PPT>0mm) percentile statewide and ZIP code-specific cutpoints exceed 0mm starting at the 1st percentile. The wet-day cutpoints always exceed the all-day cutpoints and their distributions have higher means and standard deviations. Finally, for statewide and ZIP code-specific all-day and wet-day cutpoints to be similar precipitation values (mm), they would correspond to different Xth percentile thresholds where the all-day thresholds are greater than the wet-day thresholds. For example, the cutpoint for the all-day statewide 95th (99th) percentile (19.6 mm and 43.0 mm respectively) is less than the wet-day statewide 85th (99th) percentile cutpoints (20.1 mm and 63.1 mm respectively).

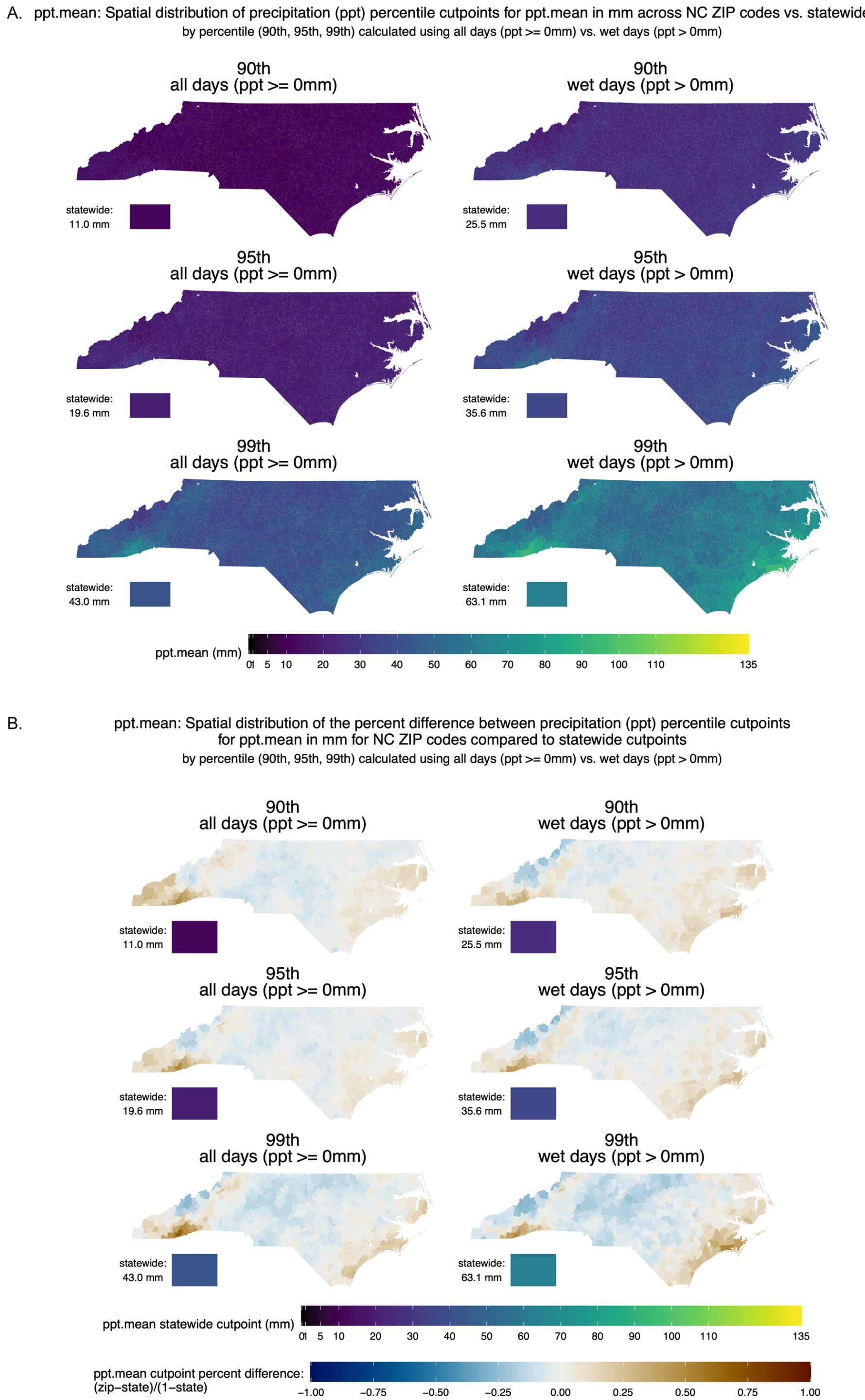


FIGURE 3. Spatial distribution of extreme precipitation (ppt) percentile cutpoints (mm) across North Carolina for spatial mean of daily total precipitation (ppt.mean) by ZIP code compared to statewide cutpoints [Panel A] for all-day (all: ppt.mean>= 0mm) [left] and wet-day percentiles (wet: ppt.mean > 0mm) [right]. Spatial distribution of the percent difference between ZIP code and statewide extreme precipitation percentile cutpoints for ppt.mean. Cutpoint percent difference calculated as (ZIP code – statewide)/(1-statewide) [Panel B]. Extreme precipitation percentiles increase from top to bottom (90th, 95th, 99th).

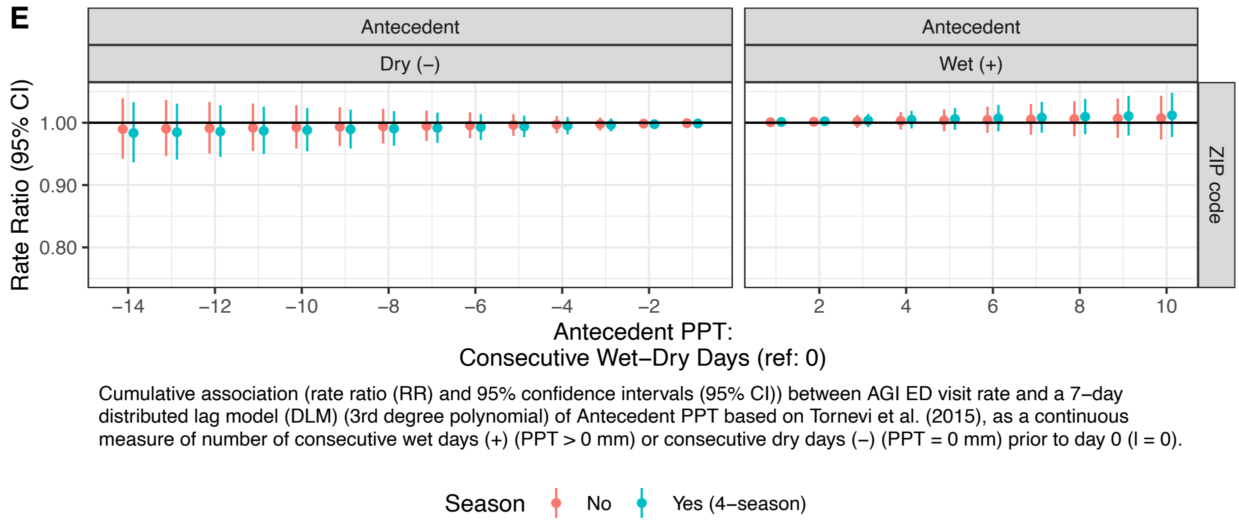


FIGURE 4. ***PANEL E:*** **Cumulative association between AGI ED visit rate and 8-day distributed lag model (DLM) of continuous wet-dry days, using a 3rd degree polynomial. This is an alternative measure for Antecedent wet-dry days, represented as a continuous measure of consecutive dry (-) and wet (+) days prior to given day (t=0; lag l = 0) compared to the categorical wet-dry day variable (FIGURE 1 Panel C). The linear trend of the cumulative association between AGI ED visit rates and wet-dry days from dry days (-) (inverse relationship) to wet days (+) positive relationship (with wide confidence intervals) is reflective of the nature of a distributed lag *linear* model (DLM), and not a distributed lag *nonlinear* model (DLNM) because the predictor-response relationship is linear. This is likely not a very good measure for this dataset because the majority of ZIP code-days lie between -1 (1 dry) to 1 (1 wet) consecutive days. Future work could explore different cutpoints defining wet days (e.g., > 1 mm) with a DLNM and alternative metrics such as a frequency index for a precipitation threshold, similar to degree days, in which the number of days that exceed a specific threshold (e.g., >1 or 7 mm) is counted for a specified period (e.g., 4-8 weeks).**

3.9. Supplementary tables

TABLE 17. Additional descriptive statistics by ZIP code-day for ambient temperature (spatial mean of Tmean, spatial min of Tmin, and spatial maximum of Tmax), mean dewpoint temperature (spatial mean of Tdmean), and spatial mean, minimum, median, and maximum of daily total precipitation (PPTmean, PPTmin, PPTmedian, and PPTmax).

TEMPERATURE (T) [1]	**Total (n)**	**Min**	**25th**	**50th**	**75th**	**Max**	**Mean**	**SD**
Mean Temperature, Tmean (°C)	2,085,710	-18.0	8.1	16.4	23.2	32.9	15.3	8.9
Min Temperature, Tmin (°C)	2,085,710	-25.1	0.9	9.6	17.2	29.0	8.8	9.3
Max Temperature, Tmax (°C)	2,085,710	-13.4	15.0	23.2	29.3	41.1	21.8	9.0
Mean Dewpoint Temperature, Tdmean (°C)	2,085,710	-25.9	1.3	10.8	17.5	27.1	9.0	9.8
ABSOLUTE PRECIPITATION (PPT)[2]	**Total (n)**	**Min**	**25th**	**50th**	**75th**	**Max**	**Mean**	**SD**
PPTmean (mm)	2,084,236	0.0	0.0	0.0	2.2	248.5	3.4	9.0
PPTmean (mm), excluding dry days (56%) (rainy days: PPTmean >0mm)	823,214	0.1	1.2	4.2	11.3	248.5	8.7	12.6
PPTmean (mm), excluding dry days (61%) (rainy days: PPTmean >0.1mm)	929,416	0.0	0.6	3.1	9.9	248.5	7.7	12.2
PPTmin (mm)	2,084,236	0.0	0.0	0.0	2.0	250.6	3.4	8.9
PPTmedian (mm)	2,084,236	0.0	0.0	0.0	0.8	238.4	2.4	7.1
PPTmax (mm)	2,084,236	0.0	0.0	0.0	4.2	279.8	5.1	12.0

[1] Mean, minimum, and maximum temperatures were calculated respectively using the spatial mean and the corresponding statistic (minimum or maximum) across ZIP code. Values shown are the spatial mean of *Tmean*, spatial minimum of *Tmin*, and spatial maximum of *Tmax*.
[2] *PPTmean*, *PPTmin*, and *PPTmax* are respectively the spatial mean, maximum, and minimum values of daily total precipitation (*PPT*) for all 1x1-km grids of PRISM data whose centroids lie within the ZIP code polygon (2013). *PPTmean* is the spatial mean of daily total precipitation (*PPT*) for all 1x1-km grids of PRISM data whose centroids lie within the ZIP code polygon (2013) and corresponds to *PPT* unless stated otherwise.

TABLE 18. Counts of consecutive wet (+) and dry (-) days following Tornevi (2013, 2015) (2008-2015). Absolute values of the Tornevi indicator variable for consecutive wet (+) or dry (-) days is displayed. Wet days were defined as having daily total precipitation greater than 0 mm. Consecutive wet and dry days prior to each day were classified into 7 categories: 1 Wet/Dry (n = 2,084,236), 2-3 Dry (n= 424), 2-3 Wet (n= 349), 4-6 Dry (n= 262), 4-6 Wet (n= 124), 7+ Dry (n= 83), and 7+ Wet (n= 26).

Antecedent Consecutive Wet (+) or Dry (-) Days (Tornevi)	**Dry (-) Days,** N = 1,154,820	**Wet (+) Days,** N = 929,416
1	1,154,051	928,917
2	256	226
3	168	123
4	120	66
5	85	36
6	57	22
7	32	14
8	24	8
9	12	3
10	9	1
11	2	0
12	2	0
13	1	0
14	1	0

TABLE 19. ***Top panel*****: Descriptive statistics for 8-week rolling sums of daily total precipitation (mm) by ZIP code-day over 1-4, 6, and 8 weeks. 8-week cumulative sums were used to calculate the precipitation cutpoints (mm) for the antecedent wet-dry tertiles (33rd and 67th percentile) shown in the bottom panel.** ***Bottom panel*****: Descriptive statistics of cutpoints (mm) for 3 definitions of antecedent 8-week cumulative precipitation aggregated by spatial unit (state, county, ZIP code) following Carlton et al. (2014) (statewide) and Lee et al. (2019) (county- and ZIP code-specific) (2008-2015). Statewide, county-specific, and ZIP code-specific antecedent rainfall is categorized into tertiles based on 8-week cumulative precipitation cutpoints (mm) for the 33rd and 67th percentiles: Dry (<33rd), Moderate (≥33rd Mod < 67th), and Wet (≥ 67th). The statewide cutpoints (33rd: 147.3 mm; 67th: 208.0 mm) fall between the median and mean cutpoints for both counties and ZIP codes. The distribution of county- and ZIP code-specific cutpoints is skewed to the right (mean>median), but have similar mean values respectively (33rd: 150.9, 148.9 mm; 67th: 211.6, 209.1 mm).**

1-week to 8-week rolling sums of daily total precipitation (mm) by ZIP code-day across NC (2008-2015)

Cumulative sums of daily total PPTmean (mm) by ZIP code over X weeks:	Total (n ZIP code-days)	Min	25th	50th	75th	Max	Mean	SD
1-week cumulative sum	2,075,392	0.0	6.0	16.9	33.5	582.2	24.1	26.6
2-week cumulative sum	2,065,074	0.0	22.6	40.1	64.1	582.5	48.0	37.2
3-week cumulative sum	2,054,756	0.0	40.7	63.4	92.1	602.5	71.6	45.4
4-week cumulative sum	2,044,438	0.2	59.9	86.3	119.0	636.6	94.9	52.4
6-week cumulative sum	2,031,909	3.9	97.9	131.6	173.0	729.7	141.8	65.2
8-week cumulative sum	2,021,591	6.4	137.5	177.3	225.8	875.6	188.4	76.5

Wet-dry tertile antecedent precipitation cutpoints (mm) derived from 8-week cumulative sums of daily total precipitation by spatial area of aggregation (statewide, county- and ZIP code-specific) and tertile (33rd and 67th percentiles)

Spatial Unit	Tertile	8-week Cutpoint (mm)	Min	25th	50th	75th	Max	Mean	SD
State	Dry (<33rd)	147.3	-	-	-	-	-	-	-
State	Wet (≥67th)	208.0	-	-	-	-	-	-	-
County	Dry (<33rd)	county-specific	132.7	142.6	146.8	154.3	221.0	150.9	14.9
County	Wet (≥67th)	county-specific	187.7	196.5	205.4	221.3	311.2	211.6	20.9
ZIP code	Dry (<33rd)	ZIP code-specific	118.3	140.8	145.7	151.8	237.8	148.9	15.1
ZIP code	Wet (≥67th)	ZIP code-specific	169.4	195.5	201.6	216.5	341.0	209.1	21.9

Note: Spatially-specific cutpoints were calculated for all 100 counties and the 737 ZIP codes in North Carolina with 2013 ZIP code polygons.

TABLE 20. Comparison of the association between the 8-day DLM for Extreme PPT only (90th, 95th, 99th PCT) (PANEL A), Antecedent PPT only and all-cause AGI ED rates (PANEL B), and the combined association of the interaction between Antecedent PPT X 95th PCT Extreme PPT (lag = 0) and all-cause AGI ED rates (PANEL C). Cumulative rate ratios (RR) (PANEL A) or rate ratios (RR) (PANELS B and C) and 95% confidence intervals (95% CI) are shown to 3 decimal places so that comparisons between the values can be made more easily. The reference categories are not shown. Rate ratios for the Antecedent only models (with and without controlling for season) show a positive association with AGI ED rates and generally lie between the values of the combined association for the Antecedent and Extreme interactions, where the association between Not Extreme X Antecedent is positive and greater than Antecedent only, but Extreme X Antecedent is inversely associated and less than the corresponding Antecedent only.

A. Model 2: EXTREME ONLY (0-7 day DLM) *(ref: Not Extreme, not shown)*		**All Days** (PPT ≥ 0 mm) RR (95% CI)			**Wet Days** (PPT > 0 mm) RR (95% CI)		
		State-wide	**County**	**ZIP code**	**State-wide**	**County**	**ZIP code**
Extreme: 90th	*No Season*	**0.998** (0.985, 1.011)		**0.986** (0.973, 0.999)	**0.924** (0.906, 0.942)		**0.929** (0.911, 0.947)
	Season	**0.994** (0.981, 1.007)		**0.982** (0.969, 0.995)	**0.918** (0.900, 0.936)		**0.922** (0.905, 0.940)
Extreme: 95th	*No Season*	**0.937** (0.920, 0.955)		**0.922** (0.906, 0.939)	**0.904** (0.879, 0.930)		**0.908** (0.884, 0.933)
	Season	**0.931** (0.914, 0.948)		**0.916** (0.899, 0.932)	**0.899** (0.874, 0.924)		**0.903** (0.878, 0.928)
Extreme: 99th	*No Season*	**0.869** (0.834, 0.905)		**0.837** (0.804, 0.871)	**0.850** (0.799, 0.905)		**0.820** (0.773, 0.871)
	Season	**0.865** (0.830, 0.901)		**0.833** (0.800, 0.867)	**0.847** (0.795, 0.902)		**0.817** (0.770, 0.867)

RR (95% CI) of the cumulative association of 90th, 95th, and 99th percentile Extreme precipitation (PPT) using a 0-7 day distributed lag model (DLM) on all-cause AGI ED rates. Cutpoints (mm) for Extreme PPT varied by all days (PPT >= 0 mm) and wet days (PPT > 0 mm), by spatial aggregation (statewide vs. ZIP code-specific), and by percentile (90th, 95th, and 99th). Not Extreme reference not shown.

B. Model 3: ANTECEDENT ONLY (wet-dry tertiles)	**Antecedent Dry** RR (95% CI)	**Antecedent Wet** RR (95% CI)

(ref: Mod, not shown)		State-wide	County	ZIP code				State-wide	County	ZIP code			
	No Season	**1.024** (1.020, 1.028)	**1.018** (1.014, 1.022)	**1.019** (1.015, 1.023)				**1.003** (0.999, 1.007)	**0.999** (0.995, 1.002)	**0.994** (0.991, 0.998)			
	Season	**1.024** (1.021, 1.028)	**1.019** (1.015, 1.022)	**1.019** (1.016, 1.023)				**1.003** (0.999, 1.007)	**0.999** (0.995, 1.003)	**0.995** (0.991, 0.998)			

RR (95% CI) of the association of wet-dry tertiles of 8-week prior Antecedent PPT (Dry, Moderate (Mod), Wet) on all-cause AGI ED rates. The spatial aggregation level for precipitation cutpoints were defined as statewide, county-specific, and ZIP code-specific. Moderate reference not shown.

C. Model 4: ANTECEDENT X 95th EXTREME (lag=0) *(ref: Mod & Not Extreme, not shown)*		**Antecedent Dry** RR (95% CI)						**Antecedent Wet** RR (95% CI)					
		All Days (PPT ≥ 0 mm)			**Wet Days** (PPT > 0 mm)			**All Days** (PPT ≥ 0 mm)			**Wet Days** (PPT > 0 mm)		
		State-wide	County	ZIP code	State-wide	County	ZIP code	State-wide	County	ZIP code	State-wide	County	ZIP code
Extreme: 95th	*No Season*	**0.975** (0.958, 0.992)		**0.978** (0.961, 0.995)	**0.943** (0.911, 0.977)		**0.956** (0.923, 0.990)	**0.983** (0.974, 0.993)		**0.984** (0.975, 0.994)	**0.979** (0.966, 0.993)		**0.985** (0.972, 0.998)
	Season	**0.970** (0.954, 0.987)		**0.973** (0.957, 0.990)	**0.938** (0.905, 0.971)		**0.951** (0.918, 0.985)	**0.984** (0.974, 0.993)		**0.984** (0.975, 0.994)	**0.980** (0.967, 0.993)		**0.985** (0.972, 0.998)
Not Extreme	*No Season*	**1.026** (1.008, 1.043)		**1.023** (1.005, 1.041)	**1.060** (1.024, 1.098)		**1.046** (1.010, 1.083)	**1.017** (1.007, 1.027)		**1.016** (1.006, 1.025)	**1.021** (1.007, 1.035)		**1.015** (1.002, 1.029)
	Season	**1.030** (1.013, 1.048)		**1.027** (1.010, 1.045)	**1.067** (1.030, 1.104)		**1.052** (1.016, 1.089)	**1.017** (1.007, 1.027)		**1.016** (1.006, 1.025)	**1.021** (1.007, 1.035)		1.015 (1.002, 1.028)

RR (95% CI) of the combined association of same-day (lag=0) 95th percentile Extreme precipitation (PPT) and wet-dry tertiles of 8-week prior Antecedent PPT (Dry, Moderate (Mod), Wet) on all-cause AGI ED rates. The spatial aggregation level for precipitation cutpoints was matched between Extreme PPT and Antecedent PPT measures (i.e., statewide and ZIP code-specific). Moderate-Extreme results and Moderate-Not Extreme reference are not shown.

4. INVESTIGATING THE EFFECT OF REGIONALITY DOMESTIC WELL WATER SUPPLY, AND INDUSTRIAL HOG OPERATIONS ON THE ASSOCIATION BETWEEN EXTREME RAINFALL AND ACUTE GASTROINTESTINAL ILLNESS IN NORTH CAROLINA (AIM 3)

4.1. Introduction

The intensity and frequency of heavy and extreme precipitation events in the United States have increased over the 20th century and are projected to continue to increase throughout this century (Easterling et al., 2017; Lall et al., 2018; Prein et al., 2017). In the southeastern U.S. though annual precipitation has decreased, flood frequencies are increasing as a combined result of increasing extreme rainfall events and sea level rise, with greater risk to coastal areas (Lall et al., 2018), and the most extreme precipitation events caused by hurricanes are likely to increase in intensity (Easterling et al., 2017). Precipitation is an important environmental driver of enteric pathogens, affecting their transport, survival, and transmission of enteric pathogens (Semenza, Herbst, et al., 2012; Semenza, Höser, et al., 2012). Heavy precipitation events can increase runoff and affect transmission of enteric pathogens in surface waters (Semenza, Herbst, et al., 2012), with possible exposure routes including direct contact with recreational waters and indirectly through drinking water supplies (surface and groundwater) and food products (plant and animal) (K. Levy et al., 2016).

Potential non-point sources of microbial contamination include reservoirs of human and zoonotic fecal organisms released from water or wastewater infrastructure such as septic system leakages, combined sewer overflows (CSOs) and sanitary sewer overflows (SSOs) (K. Levy et al., 2016), animal feces (Penakalapati et al., 2017), domesticated animals on farms via the direct application or spraying of manure, or leaks, breaches, or overflows of lagoons storing animal waste during heavy rains, flooding, or hurricanes (Quist, Fliss, et al., 2022; Quist, Holcomb, et al., 2022; Sterk et al., 2013; Wing et al., 2000; Zambrano et al., 2014). Acute gastrointestinal illness (AGI), which encompasses symptoms of diarrhea, nausea, and/or vomiting, has been associated with contaminated surface water and groundwater (Bylund et al., 2017) and proximity to industrial animal production (Zambrano et al., 2014). Though hydrometeorology plays a role in the transmission dynamics of microbial pathogens throughout the environment, the associations between precipitation and acute gastrointestinal illness have been hypothesized to vary by different effect modifiers such as pathogen etiology, study design and analysis, climatic or precipitation patterns or events (e.g., antecedent rainfall), geography (e.g., urban vs. rural), sociodemographics (e.g., age, gender, income) (Kraay et al., 2020). However, evaluating how the relationships between weather and AGI may be modified by risk factors (e.g., demographics, socioeconomics, water and sanitation infrastructure, land use, behavioral factors, etc.) has been identified as a priority research area (K. Levy et al., 2016) in order to improve the accuracy of climate change impact predictions or design appropriate adaptation strategies (Mellor et al., 2016).

In Ch. 3 (Aim 2), we observed a cumulative 1-18% decrease in AGI ED rates statewide following extreme precipitation over the prior 0-7 days, depending on the measure of precipitation, for North Carolina when controlling for time-variant factors and county. However, this statewide model is not able to identify at-risk populations or areas and models that account for important sociodemographic and environmental risk factor, including region, water infrastructure, and livestock agriculture, are needed for NC. The few studies that have examined how drinking water sources modify the relationship between rainfall and AGI have found conflicting results that vary by water supply source type or treatment (de Roos et al., 2020; Gleason & Fagliano, 2017; Teschke et al., 2010; Uejio et al., 2014). A recent North Carolina study has found that all-cause AGI ED visits were higher after heavy rain events (at least one day of 99th precipitation in the prior week) in areas with high hog CAFO exposures compared to no exposures in 2016-2019 (RR = 1.41; 95% CI: 1.19, 1.62) (Quist, Holcomb, et al., 2022), though few studies have investigated the effect modification of industrial swine operations on the association between precipitation and AGI (Febriani et al., 2010; Quist, Holcomb, et al., 2022).

The state of North Carolina offers on interesting setting to study how different factors may affect the relationship between on acute gastrointestinal illness and precipitation because of its variation in climate, geography, sociodemographics, and potential environmental exposures across the state. North Carolina is divided into three major physiographic provinces—the Blue Ridge mountains in the west, the Piedmont in central N.C., and the Coastal Plain in the east—and two main physiographic sub-regions—the Sandhills (or Upper Coastal Plain) and Tidewater region—that are distinct from the main regions in terms of vegetation and soil more than climate (Bennett & Patton, 2008, figs. 1.4, 2.3, 3.5, 7.13). To the west, the resistant metamorphic and igneous crystalline rock of the Blue Ridge mountains are separated from the Piedmont's mix of resistant metamorphic, igneous intrusions, and Triassic sedimentary layers by the Brevard and Bowens Creek Faults. To the east, the Fall Line separates the older Piedmont from the loose, unconsolidated material of the younger Coastal Plain (Bennett & Patton, 2008, figs. 1.4, 7.13, pp. 7–27). The ninth most populous state with 10.5 million people in 2021 (U.S. Census Bureau, 2022), North Carolina is experiencing rapid urbanization, yet has the largest proportion of residents in rural areas amongst the ten most populous states (20.4%; 2.15 million people) (Tippett, 2016; USDA-ERS, 2022). NC's largest urban centers, where the majority of recent growth has occurred, are located in the Piedmont and extend from Charlotte to Raleigh, collectively forming the Piedmont Urban Crescent (Trelease, 2006). Regions in NC vary by socioeconomics, race, access to public services such as municipal water supplies and hospitals, and potential sources of pollution such as industries and agriculture.

While the majority of U.S. residents are served by public water systems as their primary source of drinking water, over 40 million residents (13-14%) rely on domestic self-supply from private groundwater wells (Dieter, Maupin, et al., 2018; Maupin et al., 2014). North Carolina has one of the highest self-supplied populations, estimated at 2.4 million (24%) in 2015 (Dieter et al 2018). Private wells are distributed across the state (FIGURE 5) outside the boundaries of municipal water supplies in rural areas, as well as in peri-urban or donut-hole communities and current or historically Black communities, some of which were denied and have not yet received services since racial segregation was legal (Dewan 2005; Gilbert 2013; Johnson et al. 2004; Marsh et al. 2013). Unlike public water systems, private wells are not regulated by the Safe Drinking Water Act and, therefore, unmandated water quality monitoring and treatment are the responsibility of the well owner (US Environmental Protection Agency (EPA), 2022). Perhaps due to barriers posed by the lack of knowledge or costs of water quality testing and treatment to well water owners (Eaves et al., 2022; MacDonald Gibson & Pieper, 2017), many private wells are not regularly tested for microbial and chemical contamination and may have increased risks of exposure. Despite the large number of wells in North Carolina, 16,138 well water samples (2009-2013) (MacDonald Gibson & Pieper, 2017) and fewer than 200,000 private wells (2000-2010) (NC DHHS, 2021) were tested for contaminants. A recent study estimated that 7.3% (95% CI: 6.6, 7.9%) of all AGI-related emergency department (ED) visits in NC from 2007-2013 were attributable to microbial contamination in drinking water and 99% (29,200; 95% CI: 26,500, 31,900) of these cases were associated with domestic wells (DeFelice et al., 2016). Sources of contamination for untreated groundwater include on-site sewage or septic systems, which have been associated with endemic diarrheal illness (Borchardt et al., 2003) and waterborne disease outbreaks associated with contaminated well water (Anderson et al., 2003; Gunnarsdottir et al., 2013; Wallender et al., 2014).

In addition to having a large number of private wells, North Carolina is a top hog-producing state with 9 million hogs (USDA, 2007) that are predominantly raised on industrial hog farms known as concentrated animal feeding operations (CAFOs). CAFOs have been shown to directly and indirectly adversely affect the public and environmental health of surrounding areas through air quality and surface and groundwater quality impacts, including nutrient- and pathogen-loading from effluents off livestock farms (Burkholder et al., 2007; Hribar, 2010; U.S. GAO, 2008; USEPA-OW, 2013; Wing et al., 2000). Hog CAFOs are densely located in eastern NC and absent from the mountains, with lower hog densities in the Piedmont (FIGURE 5). These areas in eastern NC face the burden of additional health exposures including flooding and landfills (Norton et al., 2007; Stingone & Wing, 2011; Wing & Johnston, 2014).

This study expands on Ch. 3 (Aim 2) by expanding the statewide models between precipitation and AGI for North Carolina to test for effect modification at the sub-state level. We investigats the association between extreme precipitation and AGI and the influence of regional variability, residential water supply source, and industrial hog farms on this relationship in NC using time series quasi-Poisson distributed lag models.

4.2. Methods

4.2.1. Outcome

Using the same outcome dataset as Ch. 3 (Aim 2), we identified emergency department (ED) visits due to AGI amongst North Carolina (NC) residents from January 1, 2008 to September 30, 2015 using NC's statewide ED syndromic surveillance system, North Carolina Disease Event Tracking and Epidemiological Collection Tool (NC DETECT; https://ncdetect.org/) at the daily, 5-digit ZIP code level (the finest spatial resolution available) (Hakenwerth et al., 2009; Holcomb et al., 2022; Lippmann et al., 2013; Quist, Fliss, et al., 2022; Waller et al., 2011). All-cause AGI was defined by ICD-9-CM diagnostic codes for ED cases that included at least one of the following primary or secondary ICD-9-CM codes: infectious GI illness (001.xx to 009.xx), non-infectious GI illness (558.9), diarrhea (not otherwise specified) or nausea, vomiting, and diarrhea (787.91), and nausea and/or vomiting (787.01-787.03), as in recent studies (DeFelice et al., 2015; Tinker et al., 2009, 2010). In addition to ICD-9-CM diagnostic codes, we used the date of patient admission and 5-digit billing ZIP code and aggregated all-cause AGI ED visits by day and ZIP code. Because the shift from ICD-9-CM to ICD-10-CM as the diagnostic standard on October 1, 2015 coincided with reduced ED reporting rates, we restricted our analysis to ED visits occurring on or before September 30, 2015.

4.2.2. Exposure

As in Ch. 3 (Aim 2), we obtained daily, gridded (4x4-km) meteorological data from the Parameter-elevation Regressions on Independent Slopes Model (PRISM) AN81d dataset for 2006-2015 (Daly et al., 1994, 2008; PRISM Climate Group, 2004, 2016). We aggregated the daily weather data to the ZIP code-level by uniformly downscaling to 1x1-km and assigning each to the 2013 ESRI ZIP code polygon containing the centroid of each grid (Environmental Systems Research Institute (ESRI), 2013). The ZIP code-level spatial mean of each climatic variable was calculated over grids whose centroids were enclosed within a given ZIP code. ZIP code-level spatial means of daily total precipitation (PPT) in mm, transformed into an extreme precipitation index (Extreme PPT), and daily mean temperature (Tmean, °C) were used as the weather exposure variables in this study. Selecting one of the precipitation measures used in Ch. 3 (all-day ZIP code-specific 95th percentile extreme precipitation), we assessed ZIP code-specific extreme precipitation using percentiles calculated from all days in the 10-year period of 2006-2015 (including zero-precipitation days) and considered extreme PPT to be any day with precipitation at or above the 95th percentile precipitation value for that ZIP code. Extreme precipitation was evaluated over a distributed lag of 0-7 days.

4.2.3. Covariates

Estimates of the ZIP code-level percentage of population on private wells were derived from 2010 estimates of private well density by Census block from the EPA (Murray et al., 2021) (available at http://github.com/USEPA/PDW_Paper_2020). We aggregated the 2010 Census block estimates of total population and population served by private wells to 2013 ZIP code polygons and calculated the proportion of the population in each ZIP code using private wells. Tertiles of the ZIP code population percentage on private wells were used to define a categorical variable of ZIP code-level well usage for the analysis.

Hog concentrated animal feeding operation (CAFO) data were obtained from the Environmental Working Group (EWG, ewg.org) and Waterkeeper Alliance (waterkeeper.org), which has been used in prior studies (Christenson & Serre, 2017; Holcomb et al., 2022; Quist, Holcomb, et al., 2022). To generate this dataset, NC Department of Environmental Quality (NC DEQ) 2019 permit data (https://deq.nc.gov/cafo-map) was used by EWG to derive CAFO locations and head counts and then manually correct CAFO locations with satellite imagery of barn locations (Graddy et al., 2020). A permanent moratorium on new hog operations that use traditional waste management has been in place since 2007, such that we assumed the majority of hog operations in the 2019 permitting data was applicable for the study period from 2008-2015. We used the hog spatial density (number of hogs per km^2) to represent the intensity of hog CAFO exposure at the ZIP code-level. Because CAFO point locations were known, we first calculated hog spatial density within each Census block with a 5 km buffer and weighted by the block human population when calculating the mean hog spatial density in each 2013 ZIP code polygon, as described in Holcomb *et al.* (2022). Hog density was divided into three categories: no CAFOs (unexposed, zero density), low, and high hog densities (CAFOs), where high CAFOs was defined as greater than the 75^{th} percentile of non-zero densities by ZIP code.

We used 5-year American Community Survey (ACS) estimates (end year 2014) of Census block group population totals and counts by health insurance type and household income (Manson et al., 2021; Walker, 2016, 2023; Walker & Herman, 2020). As ZIP code Tabulation Areas (ZCTAs) are known to be spatiotemporally misaligned with ZIP codes (Grubesic & Matisziw, 2006; Krieger et al., 2002), we estimate ZIP code population from block group populations, as described in prior studies (Holcomb et al., 2022; Quist, Fliss, et al., 2022; Quist, Holcomb, et al., 2022). Their method uses the 2010 Census block distributions to proportionally assign the ACS block group populations by category to 2010 Census blocks and then aggregate to ESRI ZIP code polygons containing the block centroids (Quist, Fliss, et al., 2022). Using the methods of Holcomb, Quist, and Engel (2022), ZIP code median household income was estimated by fitting mean-constrained cumulative distribution functions (CDFs) to household income categories using the R package *binsmooth*. Mean household income was estimated by dividing total ZIP code income by the number of households (von Hippel et al., 2017). The same 5-year ACS data was used to estimate annual ZIP code-level populations from block group populations using the middle year of the 5-year estimates (e.g., 2015 populations were estimated from 2013-2017 5-year ACS and used for a population offset to estimate AGI ED visit rates.

A measure for rurality was estimated by assigning Census tract isolation distance scores from Doogan and colleagues (2018) (http://doogan.us/isolation/GeoIso.csv) to constituent Census blocks and calculating ZIP code-mean isolation scores weighted by 2010 population (Holcomb et al., 2022; Quist, Fliss, et al., 2022; Quist, Holcomb, et al., 2022). A categorical rurality variable was created based on national quartiles of the ZIP code-mean isolation score defined by Doogan *et al.* (2018), for which the Q2 (4.0), Q3 (4.8), and Q4 (6.1) threshold values corresponding respectively to suburban, small town, and rural categories. As there was only one ZIP code in North Carolina under the Q1 (metropolitan) cutoff score of 4.0, it was grouped with suburban to make a combined suburban/metro category.

Following DeFelice *et al.* (2015, 2016) and Vitro *et al.* (2017), we defined three North Carolina physiographic regions (Mountains, Piedmont, Coastal Plain) as a regional indicator variable to represent landform and geological differences that may affect water quality (Markewich et al. 1990) and as a proxy for salinity (Gilliam et al., 1997), which has been found to have a strong inverse relationship with fecal coliform concentrations (Mallin et al., 2000). This regional definition is also consistent with Lee *et al.* (2019)'s categorization of Georgia's physiographic regions into Piedmont (northern) and Coastal Plain (southern). A high-resolution GIS layer of physiographic region depicting the boundaries of the Blue Ridge Mountains, Piedmont, and Coast Plain was obtained by NC DEQ (https://data-ncdenr.opendata.arcgis.com/maps/ncdenr::physiographic-provinces-of-nc/explore) and used to assign ZIP codes to physiographic region based on the highest percent overlap with a region.

4.2.4. Statistical analyses

To investigate how the association between extreme precipitation and AGI may be influenced by environmental and infrastructural factors, we used time series analysis techniques to develop quasi-Poisson distributed lag models (DLM) of AGI ED rates and extreme precipitation stratified by region, percent population on private wells, and density of hog operations (CAFOs). Distributed lag models (DLM) (Gasparrini, 2011; Gasparrini et al., 2010) have been used in some time series studies of weather and diarrheal disease or acute gastrointestinal illness to estimate the cumulative association of weather exposures over multiple lags (e.g., Hall et al., 2011; Jagai et al., 2015; Phung et al., 2015). We defined the rainfall exposure as ZIP code-specific all day, 95th percentile extreme precipitation (PPT) and modeled the linear associations between extreme PPT status and AGI ED rates at lags from 0-7 days, using a cubic polynomial to constrain the distributed lag terms (*arglag*) as implemented in the R package *dlnm* (Gasparrini, 2011). Categorical variables for region (Mountains, Piedmont, Coast), well population tertiles (Low, Moderate, High), and hog density categories representing varying exposure (None, Low, High) were used as covariates in statewide and stratified models. We developed four main models: two statewide models adjusting for rurality and region (Model 0, Equation 12) and wells, CAFOs, and region (Model 1, Equation 8), and three stratified by region, private wells, and CAFOs, respectively (Models 2-4, Equation 9-11), to examine effect measure modification (EMM) by the three variables. Each stratified model was adjusted for the other two variables and all models were adjusted for same-day mean temperature, physiographic region (except in stratified model), the presence of at least one hospital within the county, log-median income, public holidays, percent without health insurance (uninsured), day of week (DOW), as well as offset of the log of ZIP code population (2014 5-year ACS). The time series models were controlled for long-term trends and seasonality using a natural cubic spline of the day-of-year six degrees of freedom interacted with an indicator term for year (N. Thomas et al., 2021). Statistical analyses were conducted in R version 4.2.0.

Equation 8. Model 1: North Carolina

$$\log(E[AGI\ ED\ visits_{it}]) \sim \beta_0 + \beta_1\ Temperature_{mean,it,l=0} + \beta_2\ \sum_{l=0,3rd\ poly}^{7} Extreme\ PPT_{95th,ZIP,all\ days,it-l} + \beta_3\ Well\ Pop\ (\%)_{tertiles,i} + \beta_4\ Hog\ CAFO\ Density_{3-cat,i} + \beta_5\ Region_i + \beta_6\ Hospital_i + \beta_7\ \log\left(Income_{median,i}\right) + \beta_8\ \log(Uninsured\ (\%)_i) + \beta_9\ Holidays_t + \beta_{10} DOW_t + \beta_{11} \mathrm{ns}(DOY_t, \mathrm{df} = 6)/Year_t + \log(Population_i)$$

Equation 9. Model 2: Stratification by Region (Mountains, Piedmont, Coast)

$$\log(E[AGI\ ED\ visits_{it}]) \sim \beta_0 + \beta_1\ Temperature_{mean,it,l=0} + \beta_2\ \sum_{l=0,3rd\ poly}^{7} Extreme\ PPT_{95th,ZIP,all\ days,it-l} + \beta_3\ Well\ Pop\ (\%)_{tertiles,i} + \beta_4\ Hog\ CAFO\ Density_{3-cat,i} + \beta_5\ Hospital_i + \beta_6\ \log\left(Income_{median,i}\right) + \beta_7\ \log(Uninsured\ (\%)_i) + \beta_8\ Holidays_t + \beta_9\ DOW_t + \beta_{10} \mathrm{ns}(DOY_t, \mathrm{df} = 6)/Year_t + \log(Population_i)$$

Equation 10. Model 3: Stratification by Private Well Population (%) Tertiles (Low, Moderate, High)

$$\log(E[AGI\ ED\ visits_{it}]) \sim \beta_0 + \beta_1\ Temperature_{mean,it,l=0} + \beta_2\ \sum_{l=0,3rd\ poly}^{7} Extreme\ PPT_{95th,ZIP,all\ days,it-l} + \beta_3\ Hog\ CAFO\ Density_{3-cat,i} + \beta_4\ Region_i + \beta_5\ Hospital_i + \beta_6\ \log(Income_{median,i}) + \beta_7\ \log(Uninsured\ (\%)_i) + \beta_8\ Holidays_t + \beta_9\ DOW_t + \beta_{10}\mathrm{ns}(DOY_t, \mathrm{df} = 6)/Year_t + \log(Population_i)$$

Equation 11. Model 4: Stratification by Hog CAFO Density Categories (None, Low, High > 75th pct)

$$\log(E[AGI\ ED\ visits_{it}]) \sim \beta_0 + \beta_1\ Temperature_{mean,it,l=0} + \beta_2\ \sum_{l=0,3rd\ poly}^{7} Extreme\ PPT_{95th,ZIP,all\ days,it-l} + \beta_3\ Well\ Pop\ (\%)_{tertiles,i} + \beta_4\ Region_i + \beta_5\ Hospital_i + \beta_6\ \log(Income_{median,i}) + \beta_7\ \log(Uninsured\ (\%)_i) + \beta_8\ Holidays_t + \beta_9\ DOW_t + \beta_{10}\mathrm{ns}(DOY_t, \mathrm{df} = 6)/Year_t + \log(Population_i)$$

Additional model specifications implemented as sensitivity analyses, including an alternate statewide model that did not adjust for CAFOs or wells and EMM analyses implemented using product-term interactions between extreme PPT and rurality, CAFOS, and wells instead of stratification, are described in the Supplementary Information.

4.2.5. Ethical statement

This study was reviewed by and received an exemption from the University of North Carolina Institutional Review Board (Study #: 15-1158) for the use of deidentified health data.

4.3. Results

As in Ch. 3, TABLE 14 displays the frequency of emergency department visits in North Carolina by AGI case definitions based on all-cause AGI, the most general case definition that was used in this analysis, in addition to alternative case definitions aggregated by pathogen type (bacterial, viral, protozoan), and selected pathogen-specific case definitions (*E. coli*, *C. difficile*, cholera). AGI ED visit frequencies are reported by year, age group, sex, health insurance type, and physiographic region. Children from the ages of 0-17 accounted for 21.9% of all-cause AGI ED visits, 65 and over for 14%, and ages 18 to 64 for 65%. The majority of ED visits were by women (64%), those with public health insurance (49%), and occurred in the Piedmont (63%). ZIP code-day-level descriptive statistics for all-cause AGI ED visits (outcome) and the meteorological exposures of mean temperature and all-day extreme precipitation are displayed in TABLE 15, which were used in both Aims 2 and 3.

The distribution of ZIP code counts by private well (left) and hog CAFO categories compared to CAFOs/wells (top), region (center), and having at least one emergency department at the county-level (bottom) for the 737 NC ZIP codes in the dataset is displayed in Supplementary TABLE 23. Most ZIP codes are located in the Piedmont (n=340; 46.1%) and Coastal Plain (n=293; 39.8%), have no hog CAFOs (n=433; 58.8%), and have access to at least one hospital with an emergency department in the county (n=658; 89.2%). Unlike hog CAFOs, private wells are distributed throughout North Carolina, such that each ZIP code has a portion of its population served by private wells as their primary drinking water source. Amongst the 247 ZIP codes in the high well population (%) tertile, the majority are located in the Piedmont (n=130; 52.6%) and Coastal Plain (n=100; 40.5%) regions and ZIP codes with no (n=116; 47.0%) or low (n=98; 39.7%) CAFO exposures. The majority of hog CAFOs are located in the Coast (n=206 of 304; 67.8%), followed by the Piedmont (n=98; 32.2%), with no hog CAFOs located in the Mountains; however, all but one of the 76 ZIP codes in the high hog CAFO category are located in the Coast.

In Ch. 3 (Aim 2), we observed a cumulative 8% decrease (CRR[10] = 0.92; 95% CI: 0.91, 0.94) in AGI rates statewide following 95th percentile or greater precipitation over 8 days when adjusting for county. In this study we observed a statewide cumulative 3% decrease (CRR = 0.97; 95% CI: 0.95, 1.00) in AGI ED rates when adjusting for wells, CAFOs, and region following greater than or equal to 95th percentile daily extreme precipitation (PPT) over 8 days (0-7 day lag) compared to non-extreme precipitation days (less than 95th percentile or no precipitation) in ZIP codes in the Piedmont with low private well populations and no hog CAFOs (TABLE 21: Model 1). In a statewide sensitivity analysis, we similarly observed a cumulative 2% decrease (0.98; 95% CI: 0.95,

[10] CRR stands for cumulative rate ratio.

1.00) in AGI ED rates after extreme precipitation over 8 days when controlling for rurality instead of private wells and hog CAFOs (TABLE 21: Model 0). All models were adjusted for same-day mean ambient temperature, county-level hospital access, the log of median income, federal holidays, percent uninsured, day-of-week, and seasonal and long-term trends. Variables for rurality, private well, and poultry CAFO density were highly correlated (not reported) and were not modeled together in the final models (0-4). In exploratory analyses, we observed that adjusting for region was the only variable that increased the cumulative rate ratio between extreme PPT and AGI ED rates and, therefore, all final models either controlled for (Models 0-1, 3-4) or stratified by (Model 2) region.

In Model 2, we examined effect measure modification (EMM) by region and observed significantly different results when stratifying by the Piedmont, Mountains, and Coastal Plain (or Coast) regions (TABLE 21). In the Piedmont, extreme precipitation was cumulatively associated with an 18% decrease (0.82; 0.79, 0.85) in AGI ED rates over 8 days compared to non-extreme days when controlling for private wells and hog CAFOs. Unlike the inverse associations between extreme precipitation and AGI ED rates observed statewide and in the Piedmont, extreme precipitation was cumulatively associated with an 18% increase (1.18; 1.07, 1.31) in AGI ED rates in the Mountains and a 19% increase (1.19; 1.14, 1.25) in the Coastal Plains.

The evidence of EMM by tertiles of private well population (%) in Model 3 (TABLE 22) on the association between extreme precipitation and AGI ED rates was inconclusive when adjusting for hog CAFOs. 95th percentile extreme precipitation (lagged 0-7 days) trended towards an inverse cumulative association with AGI ED visits in ZIP codes with low (0.98; 0.94, 1.02) and moderate (0.94; 0.90, 0.99) private well populations, but the cumulative 6% decrease in AGI ED rates following extreme precipitation was only statistically significant for the moderate well category. We observed a null cumulative association (1.00; 0.94, 1.07) between extreme precipitation and AGI ED rates in areas with high private well populations. In a statewide sensitivity analysis, we modeled the interaction between same day (0-day lag) 95th percentile extreme precipitation and private wells on AGI ED rates without controlling for hog density (Supplementary TABLE 24) and found an inverse association similar to the stratified models. On days with greater than or equal to 95th percentile precipitation, we observed statistically significant decreases in AGI ED rates by 3% (RR = 0.97; 0.96, 0.99) and 2% (0.98; 0.97, 0.99) in areas with high and moderate well populations respectively, but a statistically non-significant 1% decrease (0.99; 0.98, 1.00) in areas with low well populations, compared to non-extreme days in low well population ZIP codes. However, on days with less than 95th percentile precipitation events we observed a cumulative increase in AGI ED rates by 3% (1.03; 1.01, 1.05) and 2% (1.02; 1.01, 1.03) respectively.

We observed evidence of effect modification of extreme precipitation and AGI ED rates by CAFO categories when adjusting for region and private wells in Model 4 (TABLE 22). Extreme precipitation was associated with cumulative 15% (1.05, 1.26) and 7% (1.07; 1.02, 1.13) increases in AGI ED rates over 8 days in areas with high (>97 hogs/km^2) and low hog densities, respectively, but we observed a 13% decrease in AGI ED rates in unexposed ZIP codes with zero hog density. The positive association (15% increase) between extreme precipitation and AGI ED rates in the ZIP codes with high hog densities (Model 4), 98.7% of which are located in the Coastal Plain, is also consistent with the positive association (19% increase) in the Coast (Model 2).

4.4. Discussion

4.4.1. Major findings

We observed a 2-3% decrease in all-cause AGI ED rates following greater than or equal to 95th percentile all-day extreme precipitation over a lag of 0-7 days when adjusting for rurality and region (0.98; 0.95, 1.00) or region, private well population, and hog density (0.97; 0.95, 1.00) from 2008-2015 compared with non-extreme precipitation. This statewide inverse association between extreme precipitation and AGI ED rates is consistent in direction, though smaller in magnitude, than the 8% decrease (0.92; 0.91, 0.94) observed in the equivalent statewide model estimated in Ch. 3 when adjusting only for county-level fixed effects as opposed to ZIP-code-specific characteristics in the current study. In prior studies, the heterogeneity of the direction and statistical significance of the association between rainfall and AGI or diarrheal diseases has been noted by several systematic reviews (Guzman Herrador et al., 2015; Kraay et al., 2020; K. Levy et al., 2016). Results have varied by study characteristics including percentile threshold, covariates, spatiotemporal scale and aggregation, and data sources, such that a recent meta-analysis found non-statistically significant pooled estimates of incidence rate ratios between extreme rainfall and diarrhea with large margins of error ranging from a 36% increase (95% CI: 0.883, 2.09) at 80th percentile, 2.2% decrease (0.887, 1.08) at 90th, 2.8% decrease (0.877, 1.08) at 95th, and a null association (IRR = 1.00; 0.895, 1.12) at the 99th percentiles (Kraay et al., 2020). The few larger scale (i.e., state- or province-wide) studies on extreme precipitation and all-cause or pathogen-specific AGI in the U.S. illustrate the variation in the direction of the associations in different locations and by outcome, even without examining potential effect modifiers. In New Jersey, AGI was similarly inversely associated with 90th percentile rainfall with a 3-day lag (OR = 0.96, 95% CI: 0.92-0.99), but positively associated with 3-day average 90th percentile rainfall with a 7-day lag (OR = 1.04, 95% CI 1.01–1.08) (Gleason & Fagliano, 2017). Salmonellosis was positively associated with 90th percentile rainfall with a 1-week lag in Georgia (IRR = 1.03, 95th CI: 1.00, 1.06) (D. Lee et al., 2019) and with a 1 unit increase in 90th percentile Extreme Precipitation Threshold (EPT_{90}) metric in Maryland (IRR:1.056; 95th CI:1.035–1.078) (Jiang et al., 2015). Campylobacteriosis, however, was not associated with EPT_{90} at the state level in Maryland (Soneja, Jiang, Upperman, et al., 2016a). However, a more complex relationship between AGI and extreme precipitation is evident when accounting for potential effect modifiers. Physiographic region and hog density strongly modified the effect between extreme precipitation and AGI ED rates with changes in the direction of the association by strata, but inconclusive evidence was found for EMM by private well population.

We observed the highest AGI rates in the Blue Ridge Mountains and Coastal Plain of North Carolina, where extreme precipitation was significantly associated with approximately 18-19% increases in all-cause AGI ED rates respectively. In ZIP codes exposed to industrial hog operations—which were present only in the Piedmont and Coastal Plain regions—extreme precipitation was positively associated with AGI ED rates and rose from approximately 7% to 15% with hog density intensity, from low to high (>75th percentile; >97 hogs/km^2). Significant inverse associations between extreme precipitation and AGI were observed statewide overall (2-3% decrease in AGI rates) and for some regional, hog density, and private well strata when controlling for the remaining two unstratified covariates. The magnitude of the inverse association was highest in the Piedmont with 18% decreases in AGI, followed by areas with no hog exposures in the Piedmont, Mountains, and Coastal Plain (13% decrease), and lowest in ZIP codes with moderate private well populations (6% decrease). However, the areas with low and high well populations did not exhibit significant associations between extreme precipitation and all-cause AGI, and neither was there a consistent trend from low to high proportions of well population.

4.4.2. Effect measure modification by region

We observed substantial differences in the direction of the association between extreme precipitation and all-cause AGI rates by region compared to non-extreme precipitation when controlling for private well population, hog density, county-level hospital emergency department access, median household income, percent uninsured, holidays, and day-of-week. Similar to our study in NC, three prior statewide studies have examined and found region to be a modifier of extreme precipitation and pathogen-specific AGI: salmonellosis in the Coastal Plain and northern Georgia (GA) (D. Lee et al., 2019) and salmonellosis (Jiang et al., 2015) and campylobacteriosis

(Soneja, Jiang, Upperman, et al., 2016a) in coastal and non-coastal Maryland (MD). Though the studies used different enteric outcomes and regional definitions, it is interesting to compare the results of our NC-based study with those in GA and MD because all three states are on the U.S. Atlantic coast and share three overlapping physiographic provinces (Blue Ridge, Piedmont, Coastal Plain), with the exception of additional mountainous regions (e.g., Appalachian Plateau, Ridge and Valley) that are found in GA and/or MD. We observed increases in all-cause AGI rates in the mountains and coastal ZIP codes of NC, but decreases in the Piedmont. Similarly, campylobacteriosis was positively associated with extreme precipitation in coastal counties, but, dissimilarly, had no association in non-coastal counties of MD (Soneja, Jiang, Upperman, et al., 2016). Likewise, salmonellosis risk increased with extreme precipitation in coastal counties of both Georgia (D. Lee et al., 2019) and Maryland (Jiang et al., 2015), but had a lesser positive association in non-coastal MD and a non-significant association in Northern GA. While we observed inverse associations between all-cause AGI and 95th percentile extreme precipitation in the Piedmont, Lee and colleagues only observed inverse associations for salmonellosis (all serovars) when antecedent precipitation conditions (prior 8-week wet, moderate, or dry periods) were interacted with extreme precipitation in two instances: following a 1-week lag of non-extreme rainfall after a dry period in Coastal Plain counties and a 1-week lag of non-extreme rainfall after a wet period in Northern counties (also the only statistically significant association in the Northern counties) (D. Lee et al., 2019).

Overall, it is interesting to note that extreme precipitation was consistently positively associated with both all-cause AGI and two etiologies of bacterial AGI in coastal areas, but the direction of the associations varied in the Piedmont and mountain regions by AGI outcome (e.g., all-cause, campylobacterosis, salmonellosis) and state. In contrast to our 3-region strata for NC, the studies in GA (D. Lee et al., 2019) and MD (Jiang et al., 2015; Soneja, Jiang, Upperman, et al., 2016) both use 2-region strata: a coasta region and a single aggregated non-coastal region that included both mountainous and Piedmont regions. In the GA and MD studies, the magnitude of the non-coastal regional estimates were either attenuated positive or non-signifcant associations. Given the opposite direction (positive vs. inverse respectively), but high magnitude of the associations (18-19%) between AGI and extreme precipitation when Mountains and Piedmont were modeled separately in our study in NC, it would be interesting to further examine EMM by region by disaggregating, for example, the mountain regions from foothills (i.e., Piedmont) in other Atlantic states in future studies.

Why may the regional effects be so different? Regional variation in the association between precipitation and AGI ED visit rates in may be due to a variety of factors that affect pathogen proliferation, transportation, and exposure, including climatic, geographic, demographic, socioeconomic, and infrastructural characteristics for which we were able to control to varying degrees. In contrast to the Mountains and Coastal Plains of NC, the inverse association between AGI and 95th percentile extreme precipitation in the Piedmont is particularly notable because it may result from dilution effects instead of concentration effects (Kraay et al., 2020) and is also reflected in the statewide estimate to a lesser degree. The Piedmont represents the majority of the AGI visits in the state because it has the largest population and includes the majority of urban centers, which likely contributes to the inverse association observed statewide. Statewide and regional model differences highlight the importance of examining modification by region because different relationships between precipitation and AGI may occur in different locations as a result of underlying factors that are obscured by the size of the Piedmont's population in the Piedmont and not adequately captured in the control variables. For example, the Piedmont contains very few hog farms compared to the Coastal Plain, but more hog exposure than the mountains of western NC, whose livestock operations are primarily to raise dairy cattle and do not include hog or poultry

CAFOs (NCDEQ-DWR, 2020). In future studies, it may be worthwhile to examine effect modifiers, such as CAFO density, statewide and within relevant regions. The Coastal Plain is characterized by low water tables, highly permeable sandy soils, and exposure to seasonal hurricanes, flooding, and hog CAFOs and other industries. Additionally, rural communities in eastern North Carolina may also have increased health risks due to poorer access to healthcare and higher uninsurance rates (Hardy, 2012; North Carolina Institute of Medicine, 2018). Limited research on the relationship between climatic factors and AGI has been conducted in mountainous regions, with exceptions such as Dhimal *et al.* (2022) in Nepal and Galway *et al.* (2015) in British Columbia. In addition to precipitation, streamflow was used by Galway *et al.* (2015) as an exposure variable to compare snowfall- and rainfall-dominated regimes and may be a promising predictor to further explore, particularly for mountainous or hilly terrain. Mountainous regions are an interesting and important area for further climate-health research because of their hydrological and ecological sensitivity to climate change (Hock et al., 2022; ICIMOD, 2010) and regional variation in geological and topographical characteristics that may affect the potential transport, infiltration, and contamination of groundwater supplies by pathogens (Crane & Moore, 1984).

A factor that may be contributing to EMM by region is urban-rural geography, whose potential role in the dynamics of precipitation and diarrheal diseases has been discussed and examined in a systematic review and meta-analysis by Kraay and colleagues. The authors hypothesize as resulting from differences in access to different levels of water and wastewater infrastructure (e.g., improved versus unimproved; on-site versus piped); impervious surface coverage that increases runoff, may impact pathogen flushing and concentration-dilution dynamics following rain events, and impacts water quality (Brabec et al., 2002); and exposure to combined and sanitary sewer overflows that are a feature of many urban environments (Kraay et al., 2020). Using 2016-2019 data establish highly exposed and unexposed ZIP codes in NC, rurality was found to strongly modify the relationships between hog CAFO exposure (Quist, Holcomb, et al., 2022) and, to a lesser degree, hurricane flooding (Quist, Fliss, et al., 2022) and all-cause AGI ED rates, where higher AGI rates were observed when restricted to rural areas. The increases in all-cause AGI rates we observed in coastal areas, when controlling for private well population as a proxy for rurality, are consistent with higher AGI ED rates in rural areas observed by Quist and colleagues, and suggest the opportunity to further investigate and understand urban-rural differences by region in NC. By contrast, there was no association between precipitation and diarrhea in rural areas in Ecuador, but urban areas were associated with higher diarrheal incidence under dry antecedent precipitation conditions than wet conditions and diarrheal incidence further increased when heavy rain events followed antecedent dry periods, which the authors hypothesized may be due to the accumulation of fecal contamination during dry periods followed by the flushing of pathogens into the urban environment during heavy rains due to increased runoff with higher impervious service coverage (Deshpande et al., 2020).

4.4.3. Effect measure modification by hog CAFOs

We examined EMM by industrial hog CAFOs, which are a major issue in North Carolina due to their water quality, air quality, mental health, and environmental justice implications for communities in the vicinity (Wing & Wolf, 2000). Like regional factors, hog density was an important factor to consider because of the geographic distribution of hog CAFOs concentrated in the Coastal Plain of eastern NC and the localized effects of swine slurry application and unintentional release of hog waste from lagoons.

Rainfall intensity and runoff are factors that affect zoonotic pathogen transport in agricultural settings (Sterk et al., 2013) and heavy rainfall has been shown to transport manure-borne *Cryptosporidium* oocysts and splash fresh produce where manure is applied to soils (K. Levy et al., 2016; Sterk et al., 2013), plausible contributing to transmission pathways for water- and food-borne enteric illnesses. Compared to non-extreme precipitation, we observed increases in AGI ED rates following extreme precipitation over the prior 0-7 days, with an increasing trend from low to high (>75th percentile) hog density for ZIP codes with hog CAFOs. Conversely, AGI rates decreased following extreme precipitation in areas with no hog CAFO exposure.

The increase of AGI rates following extreme precipitation in areas with hog CAFOs supports the results of the limited number of studies that have observed positive associations with rainfall when examined the interactions between rainfall, hog CAFOs, and enteric pathogens (Eisenhauer et al., 2016; Thurston-Enriquez et al., 2005) or illnesses (Febriani et al., 2010; Quist, Holcomb, et al., 2022). Precipitation has been positively associated with increased fecal indicator organism and protozoan concentrations in runoff caused by simulated heavy rainfall from agricultural plots applied with swine slurry and cattle manure compared to control plots without manure (Thurston-Enriquez et al., 2005) and with increased *E. coli* concentrations in wells, most strongly when in the vicinity of pigs (Eisenhauer et al., 2016). In farming areas dominated by hogs, Febriani and colleagues observed that AGI was positively associated with weekly cumulative precipitation—greater than or equal to the ≥90th percentile (3–4-week lags in fall) and below the 10th percentile (4-week lag in summer)—and observed effect measure modification by farming intensity (high vs. low) and season (Febriani et al., 2010). Our results were further corroborated by a similar study by Quist, Holcomb, and colleagues (2022), who compared all-cause AGI ED visit rates in ZIP codes of high hog exposure (>75th percentile) to those without hog exposure in North Carolina in 2016-2019, during a period with two major hurricanes. The authors observed a positive association between hog exposure and AGI rates overall, with effect modification by extreme rainfall, rurality, race, and co-location of swine CAFOs with poultry CAFOs. Daily rainfall was defined as 'heavy' when above a given percentile (80th, 90th, 99th, 99.9th) for at least one day during the prior 7 days. In areas of high hog CAFO exposure restricted to days following a heavy precipitation event in the prior week, they observed similar increases in AGI rates for 80th, 90th, and

95^{th} percentile precipitation (approximately 25%), with larger increases above the 99^{th} percentile (~61 mm) (RR = 1.41; 95% CI: 1.19, 1.62) and 99.9^{th} percentile (~81 mm) (RR = 2.86; 95% CI: 2.54, 3.18) (Quist, Holcomb, et al., 2022). Considered together, the increased AGI rates at very low ($<10^{th}$) and very high ($>99.9^{th}$) percentiles suggests the possibility of non-linear effects between extreme precipitation and AGI and that more thresholds should be explored across areas with no, low, and high CAFO exposures.

When considering differences in EMM by hog density, particularly when comparing exposed and unexposed areas, it is important to consider regional and local factors that have contributed to the heterogeneous distribution across NC. The characteristics of ZIP codes with no hog CAFOs (unexposed) are very different from those with any CAFOs (high and low exposure). The inverse association between extreme precipitation and AGI in areas without CAFOs is consistent with that of the Piedmont. There are no hog CAFOs in the mountains and they have been disproportionately sited in areas highly correlated with race, income, and rurality (Son et al., 2021) with a history of environmental racism, as inhabited by most of the 18^{th}-19^{th} century enslaved Black population in NC (MacNeil, 2015). As further discussed by Quist, Holcomb, and colleagues (2022), industrial hog operations rapidly expanded in the 1990s and early 2000s in lower income, rural areas of eastern NC that have higher Black and American Indian (mainly Lumbee, Coharie, and Waccamaw Siouan) populations (Son et al., 2021; Wing & Johnston, 2014). We were unable to examine differences in the associations between extreme precipitation and AGI by race and ethnicity due to lack of data availability in NC DETECT until 2016. However, CAFOs have been strongly associated with increased AGI rates amongst American Indian, Black, Asian, and self-pay (e.g., uninsured) patients in rural areas, suggesting that it would be worthwhile to further investigate EMM by race, ethnicity, and rurality in the association between extreme precipitation and AGI.

4.4.4. Effect measure modification by domestic wells

A prior North Carolina study of AGI ED visits used a population intervention model (PIM) at the county-month level (DeFelice et al., 2016) to estimate that 99% of the approximately 7.3% (29,500) of all AGI-related ED visits per year attributable to drinking water were associated with microbial contamination of private wells compared to community water supplies. Based on this study, we hypothesized that areas with higher private well populations would be associated with increased AGI ED rates following extreme precipitation when modeled at higher spatio-temporal resolutions. Contrary to our hypothesis, the proportion of private well population in a ZIP code (low, moderate, and high tertiles) did not modify the relationship between 95th percentile extreme precipitation and AGI ED rates over the prior 0-7 days, though there was a statistically significant decrease (6%) in all-cause AGI rates following extreme precipitation only in moderate well populations. Though our studies are not directly comparable because of the different statistical models used (PIM vs. time series quasi-Poisson) and questions asked of them, it is worthwhile to note a number of additional differences. Private wells are unregulated and particularly vulnerable to groundwater contamination due to irregular testing and treating (MacDonald Gibson & Pieper, 2017). As national drinking water supply source data was last collected during the 1990 U.S. Census, data on the locations, populations of users, and water quality of private wells is limited. Possible data sources include cross-sectional estimates of locations and/or populations at different spatial scales (Dieter, Maupin, et al., 2018; T. D. Johnson et al., 2019; T. D. Johnson & Belitz, 2017; Murray et al., 2021), assumptions based on areas outside available community water supply service boundary maps, or state- or county-collected water quality data from newly constructed wells and/or voluntary requests from well owners (as used by DeFelice et al., 2016; Eaves et al., 2022). We did not model AGI risk by residents whose main drinking water source was from private wells compared to those served by community water systems, but instead the ZIP code proportion of private wells estimated from 2010 block population estimates as a proxy (Murray et al., 2021). Estimates that less than 8% of all AGI-related ED visits per year could be attributed to drinking water, use of a broader case definition of all-cause AGI compared to microbially-attributed AGI could make it difficult to accurately identify where the burden of disease due to drinking water contamination is more likely. Furthermore, it would be an ecological fallacy to attribute risk to an individual risk based on average estimates over a population in an ecological study (Gelman et al., 2001). More broadly, it may be difficult to model whether private wells increase the risk of AGI following extreme precipitation because the interactions between precipitation, drinking water source, microbial contamination, and AGI are complex; groundwater wells are not the only water sources susceptible to microbial contamination and there are multiple reservoirs of human and zoonotic fecal pathogens that can contaminate water and food supplies.

Few studies have examined how drinking water sources modify the relationship between rainfall and AGI, but the few studies that have examined this have found conflicting results. Positive associations between rainfall and AGI have been observed for surface water sources (de Roos et al., 2020; Gleason & Fagliano, 2017; Teschke et al., 2010 though statistical significance was lost after adjusting for other variables), 'other' sources defined as unmapped private wells or very small community water systems (Gleason & Fagliano, 2017), and untreated municipal water during the summer/fall (Uejio et al., 2014); and null associations for groundwater (Gleason & Fagliano, 2017), treated municipal water (Uejio et al., 2014), and private wells (Uejio et al., 2014). Quist, Fliss, and colleagues (2022) examined the modification of hurricane-associated flooding for Hurricanes Matthew (2016) and Florence (2018) by ZIP code-level well water usage (<25%, 25-50%, >50%) estimated from county-level estimates of the number of people using well water in 2015 (Dieter, Linsey, et al., 2018; Dieter, Maupin, et al., 2018) and did not find clear evidence of EMM by well water usage. With the exception of a statistically significant 43% increase (RR = 1.43; 95% CI: 1.20, 1.66) in AGI ED rates following Hurricane Matthew and a non-significant 9% decrease (RR = 0.91; 0.64, 1.18) following Hurricane Florence in areas with moderate well water usage, the authors observed non-statistically significant 10-15% increases after the hurricanes in areas with the smallest and largest proportion of residents. Though EMM by private wells for inconclusive for AGI ED rates and both extreme precipitation and hurricane flooding despite the use of higher resolution well population data in our study, more research on the interactions between wells, extreme events, and AGI is needed to explore why the magnitude of the associations was much higher and largely positive following hurricane flooding compared to the null or inverse associations for extreme precipitation. In North Carolina, private well users are heterogeneously distributed across NC and rurality and private well use are highly correlated, which has important implications for understanding exposures and outcomes, but limits our ability to comprehensively control for it or disentangle the effects of private well use and rurality with the available data. Furthermore, environmental justice issues are implicit to the consideration of populations on private wells as some “donut hole” communities have been excluded from municipal water or wastewater services from a legacy of racial discrimination (MacDonald Gibson & Pieper, 2017). AGI risk location of private wells EMM by wells may need to be examined by different regions or covariates to explore potential mediating factors such as upstream/downstream locations, built environment, topography, sociodemographics, or exposure intensity of higher or lower pathogen concentrations reflecting concentration or dilution dynamics at different times or areas.

Sources of surface and drinking water contamination from human sources (e.g., wastewater treatment plant effluent, upstream combined sewer overflows) and zoonotic sources (e.g., animal fecal shedding, livestock manure or waste from agriculture) of fecal pathogens and whose transport is affected by rainfall intensity, amongst other factors (Bylund et al., 2017; Sterk et al., 2013).

Waterborne disease outbreaks have been associated with untreated groundwater and have increased in private water systems compared to public systems (Craun et al., 2010; Wallender et al., 2014). Extreme precipitation and AGI rates have been positively associated with combined sewer overflows (CSOs) in Atlanta, GA, with higher risk in areas in of low than high poverty (Miller et al., 2022) and in regions of Massachusetts with drinking water associated to CSOs, but not areas with recreational waters exposed to CSOs or no CSO exposure (Jagai et al., 2015a). In multiple studies, septic systems have been associated with norovirus outbreaks associated with contaminated surface waters and well water, endemic diarrheal illness, microbial and fecal contamination of surface water quality (Mattioli et al., 2021) and, in a recently field study, were traced to some nearby private wells, which were found to be contaminated at least once with human fecal contamination that was also significantly associated with lagged rainfall (Murphy et al., 2020).

4.4.5. Strengths, limitations, and future research

The key strengths of this study are the use of high spatio-temporal resolution data (daily, ZIP code) over a long duration of 8 years and quasi-Poisson GLM distributed lag models in, to the best of our knowledge, one of the first statewide studies on the influence of effect measure by key characteristics (physiographic region, industrial hog density, and private well population) on precipitation and AGI in North Carolina. We were able to control for time varying trends and variables, as well as time-invariate ZIP code-level socioeconomic and healthcare access characteristics. Despite the challenges inherent to spatial private well data, we were able to make use of new Census block-level 2010 estimates of private well populations (Murray et al., 2021), which are similar but haven't been compared to earlier 2010, 2000, and 1990 estimates (T. D. Johnson et al., 2019; T. D. Johnson & Belitz, 2017). Both of these datasets offer improved spatial resolution over the use of 2015 county-level estimates by the USGS (Dieter, Linsey, et al., 2018; Dieter, Maupin, et al., 2018) in previous studies.

One of the study limitations is use of all-cause AGI instead of pathogen-specific AGI as an outcome, which serves as a broad indicator of the multiple pathogens that may lead to AGI through pathways that may be mediated by heavy rainfall in different ways. The use of ICD-9-CM and ICD-10-CM diagnosis codes are not always based on laboratory testing, which may result in records coded as gastroenteritis NOS or infectious diarrhea, and may underreport enteric illness diagnoses by specific pathogens (Scallan et al., 2018). Total AGI incidence is likely underestimated by AGI ED visits (Mead et al., 1999), which capture only a fraction of the total cases (Jones et al., 2007), because most AGI cases are self-limiting and do not require a healthcare visit (Roy et al., 2006) and, when healthcare is sought, visits are impacted by healthcare access, such as decreased access in rural areas far from an ED or for those who are not uninsured, or increased options in urban areas. Though extreme precipitation was defined as greater than or equal to 95th percentile all day (PPT >= 0mm) daily precipitation based on the exploration and results of multiple precipitation index thresholds in Ch. 3, a recent NC study on CAFOs and AGI (Quist, Holcomb, et al., 2022) found effect modification by heavy precipitation that increased above the 95th percentile and suggests sensitivity analyses to extreme precipitation thresholds may be important to test. Rurality and wells were not included in the same model due to high collinearity, but the exploration of rurality or other regional factors may be useful to elucidate whether extreme precipitation modifies the relationship between private wells and AGI in some areas of NC.

Merging and processing health, climate, environmental, infrastructural, and sociodemographic data presents multiple challenges to the researcher. We defined the spatial unit of analysis for this study using 2013 5-digit ZIP code boundaries in correspondence to the highest resolution health outcome data available from NC DETECT, as mailing addresses that would enable finer scale analyses are unavailable to protect patient privacy. We recognized the limitations of ZIP codes due to their lack of standardization and transience, considered tradeoffs of converting ZIP code to ZCTAs used in Census and ACS data (Grubesic & Matisziw, 2006), and elected to aggregate and selectively population-weight smaller spatial units (e.g., block, block groups, 4-x4-km grid) to the ZIP code-level. A well-known analytical issue in geography that can be difficult to handle in practice, the modifiable areal unit problem (MAUP) presents a challenge when using and interpreting the results of aggregated data and is often ignored by geographers and public health researchers (Manley, 2014). MAUP arises from variation in results due to problems of scale (i.e., different sizes) and aggregation (i.e., different configurations of non-continuous groups or contiguous zones) when aggregating data into artificial spatial units and grouping them within zones or spatial boundaries (e.g., administrative or ecological), but its severity cannot be ascertained in advance of an analysis (Heywood et al., 1988; Openshaw, 1984b). Related to the MAUP is the ecological fallacy is an error in reasoning that occurs when results of aggregated data are applied to make an inference about an individual in the studied group or zone (Gelman et al., 2001; Openshaw, 1984b, 1984a). For ecological study using aggregated data like ours, estimates should be interpreted as an average over an arbitrary area (e.g, ZIP code) and not as the exposure or change in risk experienced by an individual living in the area studied (Briant et al., 2010; Wakefield, 2007). Spatial units defined by administrative boundaries are also ecologically arbitrary and do not account for features of physical geography or hydroclimatology that would be relevant to better understand the linkages between rainfall and human health (Corley et al., 2018). The physiographic regions used to examine EMM in this study are not aligned with watersheds, which flow from northwest to southeast in NC, so we were not able to examine differences between upper and lower watersheds, such as those found in a large study in 35 developing countries where the lower probability of childhood diarrheal disease was associated with better upstream watershed conditions (higher tree cover) in downstream rural communities, but not statistically significant in urban communities (Herrera et al., 2017). Environmental epidemiology studies explicitly considering more ecologically and hydrologically-relevant boundaries, such as different scales of watersheds, as alternatives to administratively-based spatial units is limited and we echo the recommendations for further studies examining the role of precipitation on health by others (Corley et al., 2018; Galway et al., 2015; Leyk, Phillips, et al., 2011).

4.4.6. Conclusions

With the increasing intensity and frequency of heavy precipitation projected under climate change, investigating effect measure modification is especially important to better understand how the relationship between precipitation and AGI varies in different contexts and what risk factors or interventions may be more effective at population level.

We demonstrated complexity in the relationship between precipitation and AGI, which was modified by physiographic region and hog CAFOs. Increased AGI ED rates were observed in the mountains, coastal plains, and areas of low and high hog CAFO density of North Carolina following extreme precipitation, while there was an inverse association in the Piedmont and the effect modification by private well population was inconclusive. This study provides further insights on climate-health dynamics in the southeastern US by illustrating the importance of considering effect modification by key risk factors, especially for studies covering larger spatial scales. We expand upon prior research in North Carolina that has studied AGI risk from drinking water sources (DeFelice et al., 2015), hurricane flooding (Quist, Fliss, et al., 2022), and industrial hog and poultry farms (Quist, Holcomb, et al., 2022), with findings that support the growing evidence that AGI risk increases when communities in the vicinity of hog CAFOs are exposed to heavy rainfall. Decreasing hog CAFO density in areas that experience frequent high intensity precipitation or are prone to flooding may lead to improved health outcomes.

Expanded consideration of regional and local socioeconomic, geographic, environmental, agricultural and/or water and wastewater infrastructure factors as potential effect modifiers in future research is key to better elucidating the dynamics between precipitation and enteric pathogens and corresponding illnesses in a complex and confounded landscape. For example, comparing our results with multiple studies in Atlantic coastal states (Jiang et al., 2015; D. Lee et al., 2019; Soneja, Jiang, Upperman, et al., 2016) suggests that physiographic region is a promising effect modifier or control variable in statewide or regional studies and that it is important to disaggregate non-coastal physiographic regions (e.g., mountains, Piedmont). Further research might explore how effect modification by physiographic region varies in different areas of the world, expand AGI and climate studies in mountainous areas, and, as recommended by some prior studies (Galway et al., 2015; Leyk, Phillips, et al., 2011), compare hydrogeological or ecological boundaries, such as watersheds, as alternatives to administrative boundaries.

4.5. Tables

TABLE 21. Rate ratios (95% CI) for the association between all-cause AGI ED rates and covariates across the state of North Carolina using daily, ZIP code level data. RR for Extreme precipitation (PPT) is the cumulative association between AGI ED rates and Extreme PPT over 8 days (0-7 day lag) as modeled by a distributed lag model using a 3rd degree polynomial. Shown are NC statewide models adjusted for rurality (Model 0), wells and CAFOs (Model 1), and region-specific models (Piedmont, Mountains, Coast) (Model 2). Models 1-4 include tertiles of the percent population on well water (EPA 2010), categories of the density of hog CAFOs (None, Low ≤ 75th pct, High >75th pct), except in cases of stratification by a particular variable. All models are adjusted for same-day mean temperature, region (except in stratified model), the presence of at least one hospital within the county, median income (log-adjusted), public holidays, percent uninsured, day of week (DOW), and (not shown) long-term trends and seasonality using a natural spline of the day of year (df = 6) interacted with year, as well as offset of the log of ZIP code population (2014 5-year ACS).

	Dependent variable: All-cause AGI ED rate				
	RR (95% CI)				
	Model 0: Rurality	Model 1: Wells + CAFOs	Model 2: Piedmont	Model 2: Mountains	Model 2: Coast
Extreme PPT: 95^{th1}	0.98 (0.96, 1.01)	0.97 (0.95, 1.00)	0.82 (0.79, 0.85)	1.18 (1.07, 1.31)	1.19 (1.14, 1.25)
Tmean	1.00*** (1.00, 1.00)	1.00** (1.00, 1.00)	1.00* (1.00, 1.00)	1.00 (1.00, 1.00)	1.00*** (1.00, 1.00)
Region (ref = Piedmont)					
Mountains	0.65*** (0.64, 0.65)	0.68*** (0.67, 0.69)			
Coast	0.94*** (0.94, 0.95)	1.00 (0.99, 1.00)			
Private Wells (ref = Low)					
Mod		1.05*** (1.05, 1.05)	1.08*** (1.08, 1.09)	1.21*** (1.20, 1.22)	0.96*** (0.96, 0.97)
High		1.01*** (1.00, 1.01)	0.99*** (0.98, 1.00)	1.15*** (1.13, 1.17)	0.99*** (0.98, 0.99)
Hog CAFOs (ref = None)					
Low		0.98*** (0.98, 0.98)	0.93*** (0.92, 0.93)		1.15*** (1.15, 1.16)
High		0.99*** (0.98, 1.00)	2.48*** (2.43, 2.53)		1.16*** (1.15, 1.17)
Rurality (ref = Suburban/Metro)					
Small Town	1.14*** (1.14, 1.15)				
Rural	1.37*** (1.37, 1.38)				
Hospital (1+ in county)	1.20*** (1.19, 1.21)	1.13*** (1.12, 1.14)	1.61*** (1.59, 1.64)	1.21*** (1.19, 1.23)	1.08*** (1.06, 1.09)
log(Median Income)	0.11*** (0.09, 0.12)	0.08*** (0.07, 0.10)	0.04*** (0.02, 0.06)	0.28*** (0.21, 0.35)	0.28*** (0.25, 0.31)
Holidays	0.98*** (0.97, 0.99)	0.98*** (0.97, 0.99)	0.98*** (0.97, 0.99)	0.99 (0.95, 1.03)	0.99* (0.97, 1.00)
Uninsured (%)	1.01*** (1.01, 1.01)	1.01*** (1.01, 1.01)	0.99*** (0.99, 0.99)	1.01*** (1.01, 1.01)	1.02*** (1.02, 1.02)

	Dependent variable: All-cause AGI ED rate RR (95% CI)				
	Model 0: Rurality	Model 1: Wells + CAFOs	Model 2: Piedmont	Model 2: Mountains	Model 2: Coast
Day-of-Week (DOW) (ref = Mon)					
Tues	0.94*** (0.93, 0.94)	0.94*** (0.93, 0.94)	0.93*** (0.93, 0.94)	0.95*** (0.93, 0.96)	0.94*** (0.93, 0.95)
Wed	1.09*** (1.09, 1.10)	1.09*** (1.09, 1.10)	1.10*** (1.09, 1.10)	1.11*** (1.10, 1.13)	1.09*** (1.08, 1.10)
Thurs	1.04*** (1.04, 1.05)	1.04*** (1.04, 1.05)	1.04*** (1.04, 1.05)	1.04*** (1.02, 1.05)	1.04*** (1.04, 1.05)
Fri	1.02*** (1.02, 1.03)	1.02*** (1.02, 1.03)	1.02*** (1.02, 1.03)	1.02*** (1.01, 1.04)	1.02*** (1.01, 1.03)
Sat	0.99*** (0.99, 1.00)	0.99*** (0.99, 1.00)	0.99** (0.99, 1.00)	0.98** (0.97, 1.00)	1.00 (0.99, 1.00)
Sun	1.00* (0.99, 1.00)	1.00* (0.99, 1.00)	1.00* (0.99, 1.00)	0.99 (0.98, 1.01)	1.00 (0.99, 1.01)
Constant	1.92*** (1.84, 2.00)	9.25*** (9.18, 9.33)	195.28*** (195.18, 195.37)	0.02*** (-0.33, 0.36)	0.02*** (-0.12, 0.16)
Observations	2,057,531	2,057,531	945,959	291,928	819,644
Strata	NC (unstratified)	NC (unstratified)	Piedmont	Mountains	Coast
Rurality	Y	-	-	-	-
Wells	-	Y	Y	Y	Y
CAFOs	-	Y	Y	Y	Y
Region	Y	Y	-	-	-
Hospital	Y	Y	Y	Y	Y

Note: [1] Extreme precipitation (95^{th} percentile) is an 8-day distributed lag model (DLM) term fit with a 3^{rd} degree polynomial. The RR (95% CI) is the cumulative association between AGI ED rates and Extreme precipitation over 8 days (0-7 day lag).
Rate ratios CRR) highlighted in yellow reflect an inverse association with AGI ED rates (decrease in AGI risk) and RR in blue reflect a positive association (increase in AGI risk).

TABLE 22. Rate ratios (95% CI) for the association between all-cause AGI ED rates and covariates across the state of North Carolina using daily, ZIP code level data. RR for Extreme precipitation (PPT) is the cumulative association between AGI ED rates and Extreme PPT over 8 days (lags *l*: 0-7 days) as modeled by a distributed lag model using a 3rd degree polynomial. Models are stratified for the percentage of the ZIP code population on private wells (low, moderate, high tertiles) (Model 3) and hog CAFOs (Model 4). Models 1-4 include tertiles of the 2010 percent population on well water, categories of the density of hog CAFOs (None, Low ≤ 75th pct, High >75th pct), except in cases of stratification by a particular variable. All models are adjusted for same-day mean temperature, region (except in stratified model), the presence of at least one hospital within the county, median income (log-adjusted), public holidays, percent uninsured, day of week (DOW), and (not shown) long-term trends and seasonality using a natural spline of the day of year (df = 6) interacted with year, as well as offset of the log of ZIP code population (2014 5-year ACS).

	Dependent variable: All-cause AGI RR (95% CI)					
	Model 3: Low Wells	Model 3: Mod Wells	Model 3: High Wells	Model 4: No CAFOs	Model 4: Low CAFOs	Model 4: High CAFOs
Extreme PPT: 95th[1]	0.98 (0.94, 1.02)	0.94 (0.90, 0.99)	1.00 (0.94, 1.07)	0.87 (0.84, 0.90)	1.07 (1.02, 1.13)	1.15 (1.05, 1.26)
Tmean	1.00 (1.00, 1.00)	1.00** (1.00, 1.00)	1.00*** (1.00, 1.00)	1.00 (1.00, 1.00)	1.00*** (1.00, 1.00)	1.00*** (1.00, 1.00)
Region (ref = Piedmont)						
Mountains	0.63*** (0.62, 0.64)	0.66*** (0.65, 0.67)	0.75*** (0.73, 0.77)	0.65*** (0.65, 0.66)		
Coast	0.91*** (0.90, 0.92)	1.00 (0.99, 1.01)	1.14*** (1.13, 1.15)	0.94*** (0.93, 0.94)	1.05*** (1.05, 1.06)	0.44*** (0.39, 0.48)
Private Wells (ref = Low)						
Mod				1.16*** (1.15, 1.16)	0.90*** (0.90, 0.91)	0.90*** (0.89, 0.92)
High				1.12*** (1.11, 1.12)	0.86*** (0.85, 0.87)	1.03*** (1.02, 1.05)
Hog CAFOs (ref = None)						
Low	1.17*** (1.17, 1.18)	0.88*** (0.87, 0.88)	0.86*** (0.86, 0.87)			
High	1.26*** (1.24, 1.27)	0.80*** (0.79, 0.81)	0.96*** (0.95, 0.97)			
Rurality (ref = Suburban/Metro)						
Small Town						
Rural						
Hospital (1+ in county)	1.07*** (1.06, 1.09)	1.09*** (1.08, 1.11)	1.40*** (1.38, 1.42)	1.28*** (1.27, 1.30)	1.10*** (1.08, 1.11)	1.05*** (1.03, 1.07)
log(Median Income)	0.08*** (0.06, 0.11)	0.06*** (0.04, 0.09)	0.11*** (0.08, 0.15)	0.07*** (0.05, 0.08)	0.15*** (0.12, 0.18)	0.26*** (0.19, 0.32)
Holidays	0.97*** (0.96, 0.98)	0.98* (0.97, 1.00)	1.00 (0.98, 1.02)	0.98*** (0.96, 0.99)	0.98** (0.97, 1.00)	1.01 (0.98, 1.04)
Uninsured (%)	1.01*** (1.01, 1.01)	1.00*** (1.00, 1.00)	1.00*** (1.00, 1.01)	1.01*** (1.01, 1.01)	1.01*** (1.01, 1.01)	0.99*** (0.99, 1.00)

	Dependent variable: All-cause AGI RR (95% CI)					
	Model 3: Low Wells	Model 3: Mod Wells	Model 3: High Wells	Model 4: No CAFOs	Model 4: Low CAFOs	Model 4: High CAFOs
Day-of-Week (DOW) (ref = Mon)						
Tues	0.93*** (0.92, 0.93)	0.94*** (0.94, 0.95)	0.95*** (0.94, 0.96)	0.94*** (0.94, 0.95)	0.93*** (0.92, 0.94)	0.94*** (0.93, 0.95)
Wed	1.09*** (1.08, 1.10)	1.10*** (1.10, 1.11)	1.09*** (1.08, 1.10)	1.10*** (1.10, 1.11)	1.09*** (1.08, 1.09)	1.09*** (1.07, 1.10)
Thurs	1.04*** (1.04, 1.05)	1.04*** (1.04, 1.05)	1.04*** (1.03, 1.05)	1.04*** (1.03, 1.04)	1.05*** (1.04, 1.05)	1.05*** (1.04, 1.06)
Fri	1.03*** (1.02, 1.03)	1.02*** (1.01, 1.02)	1.02*** (1.01, 1.03)	1.02*** (1.02, 1.03)	1.02*** (1.02, 1.03)	1.02*** (1.01, 1.03)
Sat	0.99* (0.99, 1.00)	0.99*** (0.98, 1.00)	1.00 (0.99, 1.01)	0.99** (0.99, 1.00)	0.99** (0.99, 1.00)	1.00 (0.99, 1.01)
Sun	1.00 (0.99, 1.00)	0.99*** (0.98, 1.00)	1.00 (0.99, 1.01)	1.00 (0.99, 1.00)	0.99* (0.99, 1.00)	1.00 (0.99, 1.01)
Constant	8.98*** (8.86, 9.09)	40.73*** (40.60, 40.85)	1.66*** (1.49, 1.83)	23.37*** (23.28, 23.46)	0.55*** (0.40, 0.70)	0.16*** (-0.14, 0.46)
Observations	673,680	693,329	690,522	1,204,203	639,996	213,332
Strata	Low Wells	Mod Wells	High Wells	No CAFOs	Low CAFOS	High CAFOS
Rurality	-	-	-	-	-	-
Wells	-	-	-	Y	Y	Y
CAFOs	Y	Y	Y	-	-	-
Region	Y	Y	Y	Y	Y	Y
Hospital	Y	Y	Y	Y	Y	Y

Note: [1] Extreme precipitation (95th percentile) is an 8-day distributed lag model (DLM) term fit with a 3rd degree polynomial. The RR (95% CI) is the cumulative association between AGI ED rates and Extreme precipitation over 8 days (0-7 day lag).
Rate ratios CRR) highlighted in yellow reflect an inverse association with AGI ED rates (decrease in AGI risk) and RR in blue reflect a positive association (increase in AGI risk).
*p**p***p<0.01; p-values are not shown for the Extreme precipitation (PPT).

4.6. Figures

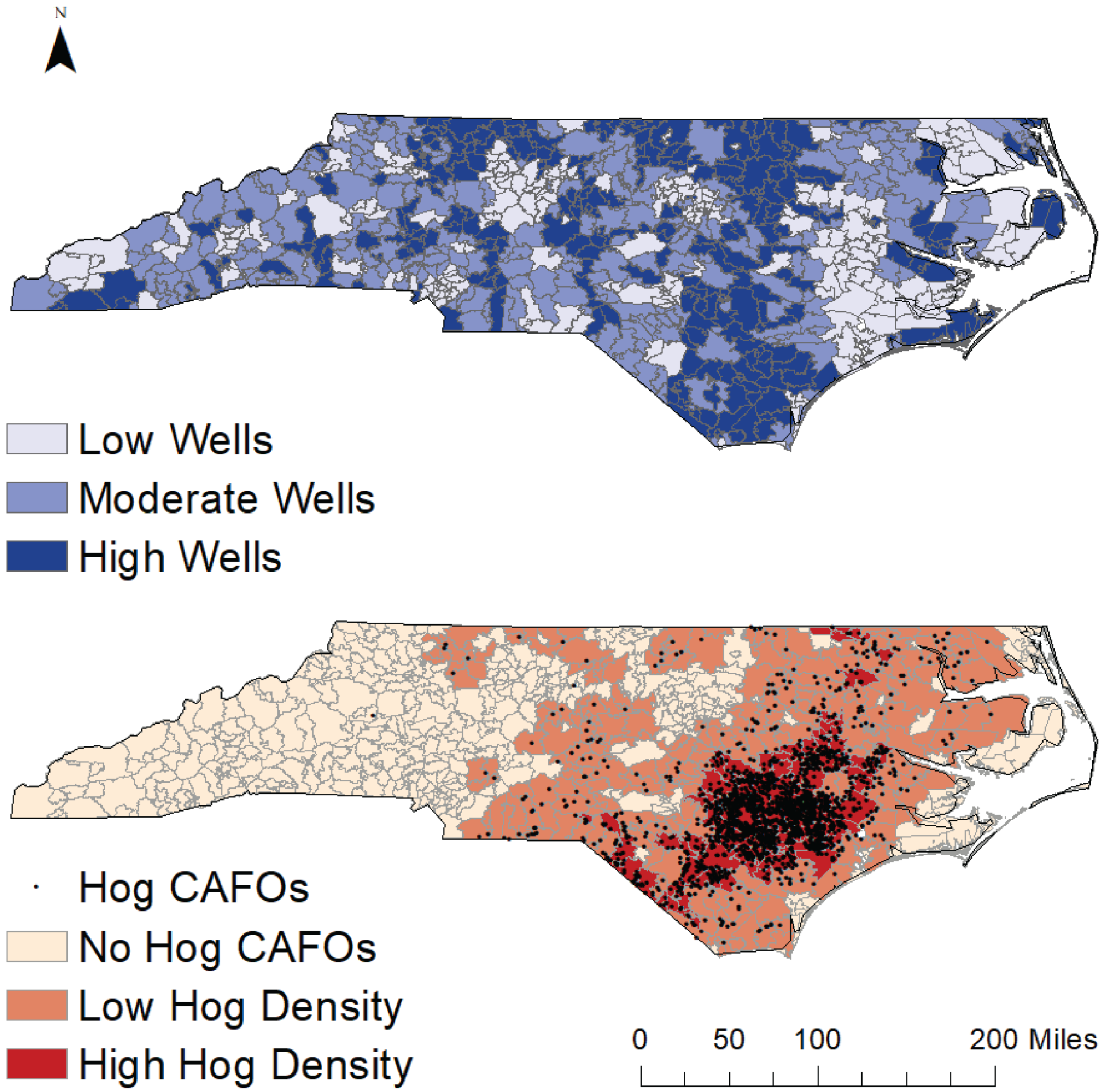


FIGURE 5. ***Top*: Spatial distribution of tertiles of private well population (%) (low, moderate, high) across NC ZIP codes (2013), aggregated from 2010 block group estimates** (Murray et al., 2021)**.** ***Bottom*: Locations of hog concentrated animal feeding operation (CAFO) locations (black) and spatial distribution of hog density categories (no CAFOS, low, high >75th percentile) across NC ZIP codes. Hog concentrated animal feeding operation (CAFO) locations and hog densities are based on 2019 NC Department of Environmental Quality swine permit data estimates available from the Environmental Working Group and Waterkeepers Alliance (Environmental Working Group (EWG) & Waterkeeper Alliance, 2016). *Figure credit*: Arbor Quist.**

4.7. Supplementary tables and equations

Equation 12. Model 0: Alternative specification of North Carolina statewide model using rurality instead of well population tertiles and CAFO density categories

$$\log(E[AGI\ ED\ visits_{it}]) \sim \beta_0 + \beta_1\ Temperature_{mean,it,l=0} + \beta_2\ \sum_{l=0,3rd\ poly}^{7} Extreme\ PPT_{95th,ZIP,all\ days,it-l} +$$

$$\beta_3\ Well\ Pop\ (\%)_{tertiles,i} + \beta_4\ Region_i + \beta_5\ Hospital_i + \beta_6\ \log\left(Income_{median,i}\right) +$$

$$\beta_7\ \log(Uninsured\ (\%)_i) + \beta_8\ Holidays_t + \beta_9\ DOW_t + \beta_{10}\mathrm{ns}(DOY_t, \mathrm{df} = 6)/Year_t + \log(Population_i)$$

TABLE 23. Counts of ZIP codes by attribute categories: private wells (low, moderate, high), hog concentrated animal feeding operation (CAFO) categories, regions (Mountains, Piedmont, Coast), and at least one hospital at the county-level with an emergency department (ED) reporting surveillance data to NC DETECT.

		Private Wells[1] (N = ZIP codes)				**Hog CAFOs[2]** (N = ZIP codes)					
		Low	*Mod*	*High*	*N*	*None*	*Low*	*High*	*N*		
CAFOs	*None*	179	138	116	*433*	179	51	13	*243*	*Low*	**Wells**
	Low	51	79	98	*228*	138	79	30	*247*	*Mod*	
	High	13	30	33	*76*	116	98	33	*247*	*High*	
Region	*Mountains*	30	57	17	*104*	104	0	0	*104*	*Mountains*	**Region**
	Piedmont	114	96	130	*340*	242	97	1	*340*	*Piedmont*	
	Coast	99	94	100	*293*	87	131	75	*293*	*Coast*	
ED in County	*No Hospital*	32	24	23	*79*	49	21	9	*79*	*No Hospital*	**ED in County**
	1+ Hospital	211	223	224	*658*	384	207	67	*658*	*1+ Hospital*	
	N	*243*	*247*	*247*	**737**	*433*	*228*	*76*	**737**	*N*	

[1] ZIP code-level private well population (%) estimates were categorized into tertiles with cutpoints for moderate and high respectively (34.8%, 63.3%).
[2] Hog CAFOs exposures were categorized based on the 75th percentile of population weighted hog density (97 hogs/km^2) by ZIP code, if any CAFOs were present in the ZIP code. The three hog CAFO exposure categories were: No CAFOS, Low CAFOs (≤75th), High CAFOS (>75th).

TABLE 24. Rate ratios and 95% CI from effect measure modification (EMM) of 95th percentile extreme precipitation (lag *l*=0) (Extreme/Not Extreme) and AGI by well population tertiles (High, Moderate, Low) statewide, controlling for region, hospital, percent uninsured, holidays, DOW, and long-term trends and seasonality. This analysis used an interaction term between extreme precipitation and well population tertiles without adjusting for CAFO density categories. High, moderate, or low private well populations are associated with decreases in AGI ED rates following a same-day 95th percentile extreme precipitation event, while high and moderate well populations are associated with increases in AGI ED rates following non-extreme precipitation, compared to non-extreme events in low well population areas. In Model 3, stratification by well population tertiles was selected over EMM because the interaction between wells and extreme precipitation appears to vary between regions.

95th percentile Extreme PPT (l=0)	Wells	RR (95% C)
Extreme	High	0.97 (0.96, 0.99)
	Mod	0.98 (0.97, 0.99)
	Low	0.99 (0.98, 1.00)
Not Extreme	High	1.03 (1.01, 1.05)
	Mod	1.02 (1.01, 1.03)
	Low	REF

5. CONCLUSIONS

5.1. Summary of findings

This dissertation contributes to the climate-health literature through an original study of the relationship between acute gastrointestinal illness (AGI) emergency department (ED) visit rates and weather, particularly precipitation, in North Carolina.

We investigated three aims in this work. First, we conducted a systematic review of the methodologies of 98 recent studies on the association between diarrheal diseases and weather. From them we identified the following lines of inquiry that we gauged worth further investigation: spatiotemporal aggregation and boundaries, exposure measures for precipitation (e.g., absolute, extreme, and antecedent precipitation), and effect modification. We then selected precipitation measures and effect modification for assessment in Aims 2 and 3.

In Aim 2, we investigated the association between multiple measures of precipitation (absolute, extreme, and antecedent) and AGI in NC from 2008-2015. Specifically, we used an ecological, time series study to estimate AGI risk associated with short-term changes in weather across NC, using high-resolution data at the daily, ZIP code-level. We observed an inverse relationship between both absolute and extreme measures of precipitation and AGI ED visits statewide. Absolute precipitation was cumulatively associated with 1% decrease in AGI ED rates at 10 mm of daily precipitation over 7 days (CRR = 0.99; 95% CI: 0.99, 1.00), 3% decrease at 40mm (0.97; 0.95, 0.98), and 6% decrease at 80mm (0.94; 0.90, 0.97).

We conducted sensitivity analyses to the definition of extreme precipitation using four definitions of extreme precipitation indices, varied by spatial references area (statewide vs. ZIP code specific) and inclusion or exclusion of non-precipitation days (all-day vs. wet-day), with three extreme precipitation thresholds (90th, 95th, 99th percentile) for each. Depending on the definition of extreme precipitation, we observed a cumulative 1%-18% decrease in AGI ED rates from extreme precipitation over 7 days. The inverse relationship between extreme precipitation and AGI ED rates was consistent for 90th, 95th, and 99th percentile extreme precipitation across definitions, with a stronger effect for wet-day extreme precipitation than all-day for a given threshold. We observed that the relationship between AGI ED visit rates and extreme precipitation was not sensitive 4-season adjustment nor to spatial reference area (statewide vs. ZIP code) over which the precipitation index was defined, though the cumulative rate ratio was stronger for ZIP code than statewide percentiles at the 99th percentile. When defining precipitation as the tertiles of total 8-week antecedent precipitation (wet, moderate, dry), we observed a cumulative increase of approximately 2% in AGI ED rates associated with dry periods compared to moderate rainfall, robust to the tertiles being defined at the statewide- (CRR = 1.02; 95% CI: 1.02, 1.03), county- (1.02; 1.01, 1.02), or ZIP code-level (1.02; 1.02; 1.02). Prior wet periods were not statistically significantly associated with AGI ED visit rates. Effect modification of same-day 95th percentile precipitation by antecedent precipitation showed a 2-6% increase in AGI ED rates for non-extreme precipitation following wet or dry periods and a 1-6% decrease in AGI ED rates for extreme precipitation compared to moderate periods of antecedent precipitation.

The concentration-dilution hypothesis suggests that heavy rain events may flush fecal pathogens and material that have accumulated in the environment during a dry period into surface waters, increasing AGI incidence (Kraay et al., 2020). Conversely, heavy rain events following rainy periods may dilute the concentration of fecal pathogens in surface waters, decreasing AGI incidence (Kraay et al., 2020). Statewide, our results for an inverse relationship between AGI and absolute or extreme precipitation are consistent with a dilution effect, while the positive relationship between

AGI and dry periods suggests a concentration effect. However, the empirical results of effect modification of 95th percentile extreme precipitation by antecedent precipitation are more difficult to interpret in light of the concentration-dilution hypothesis and warrant further investigation, particularly at lower thresholds of heavy precipitation (<90th percentile) and different lags. We observed increases in AGI with all non-extreme precipitation, following wet or dry periods, and the trend between extreme precipitation and AGI ED rates would suggest that the relationship between extreme precipitation and AGI rates may become positive at extreme precipitation thresholds lower than the 90th percentile of all-day precipitation.

In Aim 3, we evaluated whether the statewide associations observed in Aim 2, which used county-level fixed effects to control for geographic variation, were robust to effect measure modification by region, hog CAFO exposure, and private well population., we found that the *statewide* inverse relationship between 95th percentile extreme precipitation and AGI ED visits rates was robust across multiple models whether controlling for (a) county, (b) rurality and region, or (c) private wells, hog CAFO density, and region compared to non-extreme precipitation days over a lag of 0-7 days.

However, a different story emerged from the stratified models. We observed a strong decrease in AGI ED visit rates in the Piedmont associated with 95th percentile extreme precipitation, but strong increases in AGI ED rates in the Blue Ridge mountains and coastal plans of NC. We found a strong inverse association between AGI ED rates and extreme precipitation in areas with no hog CAFOs, compared to a positive association in areas with hog CAFO exposure, increasing with higher hog densities. EMM by private wells was inconclusive.

Together, Aims 2 and 3 yielded statewide and regional estimates of the association between AGI ED rates and precipitation for the state of NC and provided comparative analyses of precipitation exposure (Aim 2) and effect modification (Aim 3) that may be considered as a methodological case study of the sensitivity of the association to different model specifications. This dissertation underscores the importance of considering effect measure modification, especially with larger scale models like ours, to account for spatial and population variations. Overall, having a better understanding of the sensitivity of the association between AGI and weather to model choices is useful to researchers whether the results are sensitive or robust to those choices.

5.2. Strengths and limitations

5.2.1. Strengths

To the best of our knowledge, Aims 2 and 3 together are the first statewide time series analyses to thoroughly model the relationship between AGI and weather across the state of North Carolina, using high spatio-temporal resolution data (daily, ZIP code) over a long duration of 8 years. Aim 2 drew upon available climate science literature (Schär et al., 2016), which highlighted differences between and sensitivities of heavy precipitation indices (e.g., all-day, wet-day, frequency-based indices), to inform a novel model comparison of multiple precipitation exposure definitions of different types. Some recent studies (K. F. Bush et al., 2014b; Carlton et al., 2014; Chhetri et al., 2017; Graydon et al., 2022; D. Lee et al., 2019; Mertens et al., 2019; Tornevi et al., 2013, 2015) have compared two measures of rainfall (e.g., extreme precipitation vs. antecedent dry/wet periods) or tested the sensitivity of the precipitation measures to different cut points, percentile thresholds or lags. However, to the best of our knowledge, no AGI-weather or climate-health studies have conducted a systematic comparison of different precipitation exposures and their effects on the estimates of the associations between the health outcome and weather exposure. We were also able to assess potential effect modifiers (region, CAFOs, private wells), which have been identified as risk factors for microbial water quality or AGI in prior studies, on the relationship between AGI and precipitation. Despite the challenges inherent to spatial private well data, we were able to use new Census block-level 2010 estimates of private well populations (Murray et al., 2021).

5.2.2. Limitations

A number of study limitations arise due to the nature of the aggregated outcome and exposure data available for this study and their different sources. First, associations from ecological study designs, like this one, are subject to the 'ecological fallacy' because they use aggregate outcome and exposure data and do not imply causation at the individual level (Piantadosi et al., 1988). Results should, therefore, be interpreted with caution.

Second, there is measurement error in outcome variables and independent variables. Outcome data are based on ICD-9-CM codes and not limited to laboratory-confirmed diagnoses, making it more challenging to disaggregate by etiology (diarrheal pathogen). As discussed in Ch. 2, etiological identification is important because different pathogens may have different seasonal patterns, relationships with weather exposures, and pathways for infection. The all-cause AGI outcome variable used in Aims 2 and 3 is subject to errors of inclusion and exclusion as it largely composed on symptomatic diarrhea, which may be a mix of waterborne illnesses, foodborne illnesses, and other causes. Non-linearities and longer lag periods could also be further explored as precipitation and temperature were modeled as having linear relationships between outcome and exposure primarily over 0-7 day lags. There may also be measurement or misclassification error in independent (exposure and control variables) because weather, population, private well, and CAFO data were obtained from different sources, were often spatially aggregated, and, with exception of data directly from NC DETECT, were not linked to individual cases.

Due to data privacy limitations, the spatial location (billing address) of outcome data could only be known to the ZIP code level, so more specific information could not be related through geocoded addresses. ZIP codes associated with P.O. boxes without a matching spatial polygon were also excluded from the outcome data. Additionally, the NC DETECT data reports the billing address ZIP codes of ED patients, but census and ACS data is available only at the ZCTA, block group, and/or block level. ZCTAs and ZIP codes do not cover the same spatial areas and are known to introduce error when attempting to merge data between them (Grubesic & Matisziw, 2006; Krieger et al., 2002). To minimize this error, block group (downscaled to the block level) or block data was aggregated up to the ZIP code-level instead of attempting to cross-walk ZCTAs and ZIP codes, but this method is still subject to errors of inclusion and exclusion.

Third, the NC DETECT database provides public health surveillance data, which are subject to underreporting and, therefore, underestimate a population's true burden of disease (Gibbons et al., 2014). The database is also limited to emergency departments (excluding healthcare offices and hospitals). The underreporting of AGI cases would introduce bias into our results if variation in antecedent weather influenced the likelihood of AGI reporting; however, this is not likely (Galway et al., 2015). Furthermore, there may be selection bias in the emergency department cases because all populations in North Carolina may not be equally likely to visit an ED, potentially due to differences in healthcare access (physical or financial), or severity of illness, amongst other reasons. Lastly, in 2013-2014 there were NC DETECT reporting errors in Charlotte/Mecklenburg County, the largest county in NC, that are known to have caused underreporting.

Fourth, ecological population-based time series analysis has limitations. Time series analysis associates time varying exposures to time-varying event counts and implicitly controls for time-invariant individual-level confounders when it compares a population against itself (Dominici et al., 2003). The analysis is unable to distinguish between competing hydrologic causal pathways for AGI (Uejio et al., 2014), is subject to misclassification of environmental and demographic variables that were based on ZIP code, and is limited by the sample size in areas with lower populations and rates of AGI.

Fifth, spatio-temporal aggregation may mask relationships at smaller spatial and temporal scales. For example, weekly aggregation may mask relationships at daily time scales (should daily analyses prove infeasible due to lack of power in smaller NC counties). Use of administrative boundaries like counties and ZIP codes may also mask the actual relationship between climatic and health factors in the natural socio-ecological environment. Depending on data availability and exploratory analyses this study may or may not be comparable with watershed-level models. Though the time series models in Aims 2 and 3 were conducted at the daily, ZIP code-level, all ZIP codes were included in the statewide and stratified models, and there may have been spatial patterns in the residuals that were not accounted for and results may have been different at smaller scales.

5.3. Implications

This dissertation has underscored the importance of considering different precipitation measures, the hazards of large scale or aggregated models, and the importance of considering potential at-risk populations through effect modification. The NC DETECT dataset covers the entire state of North Carolina, which allows for the comparison of potentially heterogeneous effects of antecedent weather across different populations and settings. We identified the mountain, coastal, and CAFO-exposed areas of NC as at higher risk of AGI following extreme precipitation. The shift in direction of the association between extreme precipitation and AGI ED rates with regionality suggests that the interaction between extreme and antecedent may vary by region and should be explored further. Though we did not find private well population to modify the effect of extreme precipitation on AGI with the available data and specification, Aim 3 contributes to the recent evidence for water infrastructure (private wells) acting as effect modifiers of the relationship between weather and AGI (e.g., Carlton et al., 2014; Gleason & Fagliano, 2017) and further research is merited. While well data is limited because private wells are not regulated and water testing voluntary, private wells remain a known exposure route for possible microbial (and inorganic) contamination under conducive conditions. As areas located in the mountains and near CAFOs were associated with increased AGI ED rates, a portion of the exposure may be due to well contamination following heavy rains. If resources are limited for statewide private drinking water well testing, these types of research efforts can inform resource allocation in combination with other evidence. For example, this research may suggest that well water testing may be particularly important in mountainous, coastal, and hog CAFO-exposed areas.

Regionality and consideration of local socioeconomic, environmental, and/or infrastructural factors may be key to better elucidating the dynamics between precipitation and AGI in a complex and confounded system. In light of the results of this dissertation, state decision-makers should be aware of the variability of the relationship between extreme precipitation and AGI ED rates across the state, not captured in the statewide models, and the need for cautious interpretation of the non-causal association, especially without further understanding of the potential mechanisms at play in the relationship between precipitation and AGI. However, our observations of increased AGI ED rates associated with 95th percentile extreme precipitation in areas exposed to hog CAFOs—higher in areas with greater hog densities—does add to the growing body of evidence of the adverse health effects of hog CAFOs and the role of rainfall and flooding in hog waste lagoon leaks, breaches, and overflows that can spread fecal contamination into the environment. Precautions may be taken (a) to minimize exposure from compromised hog waste lagoons (e.g., relocation of those most likely to be affected by heavy precipitation or flooding) and (b) to improve residential water quality (e.g., education on household water treatment or boiling after heavy rainfall or flooding, or well water testing programs in CAFO-exposed areas) (Quist, Fliss, et al., 2022).

5.4. Future Research

There are multiple ways that future research could build upon the methods and results of this dissertation.

First, future research is merited into the effects of spatio-temporal aggregation, boundaries, and scale on the relationship between AGI and weather to examine the modifiable areal unit problem (MAUP) and modifiable temporal unit problem (MTUP) in the context of climate-health studies. For example, analyses for NC or other states could be expanded to compare different spatial areas based on municipal boundaries (e.g., ZIP code vs. county-level) with more hydrologically-relevant models based on watershed boundaries. If the results from the watershed analysis are promising, the study could be extended by developing raster cells that define the stream network of the hydrologic regime within each hydrologic or watershed region, similar to Leyk *et al.* (2011c). Furthermore, time series models could be compared using different time steps (e.g., daily, weekly, monthly) to test the sensitivity of the AGI-weather association to temporal aggregation, expanding upon similar work by Alarcon Falconi *et al.* (2020) for seasonal infections. Studying the effects of temporal aggregation could be complemented by developing and comparing the effects of different exposure metrics (e.g., extreme and antecedent precipitation) appropriate to each temporal unit of analysis. For monthly or longer time steps, frequency-based precipitation indices and the extreme temperature/precipitation threshold metrics (ETT, EPT) could be validated with additional datasets and more precipitation thresholds (Jiang et al., 2015; Soneja, Jiang, Fisher, et al., 2016; Soneja, Jiang, Upperman, et al., 2016; Upperman et al., 2015).

Second, future research could compare and, as needed, develop current and improved measures of antecedent precipitation at different levels of temporal aggregation (e.g., daily, weekly, month) and spatial scales. This work may also include comparisons of effect measure modification by interaction terms versus stratification, and how to best to handle EMM when using distributed lag (non-linear) models, which cannot be interacted with other terms. For North Carolina, precipitation thresholds below the 90th percentile should be assessed to see if the effect becomes positive, particularly statewide and in the Piedmont.

Third, effect measure modification can be further explored and should be considered when feasible to help identify populations at risk, to inform study similarities and differences, and to provide insights into adaptation strategies for climate change. For example, one could further develop regional models that are better customized to potential local exposures and risk factors (with and without additional covariates). Furthermore, since it is difficult to obtain accurate and recent private well data, one could compare different sources of water and sewer data, including the two recent well datasets (T. D. Johnson et al., 2019; Murray et al., 2021) versus using public water supply boundaries. Lastly, extreme and antecedent precipitation could be examined in hydrologically-relevant models (e.g., watersheds) or considered in conjunction with or contrast to runoff.

Fourth, future work could examine the effect of the source of health outcome data on the models by comparing models using ED data, as in this study, with internet search volume related to diarrheal and AGI symptoms using a method similar to that employed by Shortridge and Guikema (2014). An even more interesting outcome data source for gastrointestinal illness would be over-the-counter (OTC) medication sales for the anti-diarrheal drug, but that data may be difficult or expensive to obtain. Das *et al.* (2005) compared OTC medication sales for gastrointestinal illness and influenza-like illness with ED visits in New York City and found that the correlation between ED visits and antidiarrheal medication sales varied by season and illness, but suggested that OTC syndromic surveillance may serve as an early indicator of disease outbreaks in addition to ED surveillance systems.

Fifth, this work could be expanded using pathogen-specific outcomes derived from ICD-9 or ICD-10 codes in comparison to all-cause AGI.

Finally, the relative risk estimates of AGI in relation to temperature, precipitation, and covariates resulting from this study could be used as a basis to model climate change projections of AGI in North Carolina into the future. A similar progression was recently modeled by Chhetri and colleagues (2019) who used the results from Chhetri *et al.* (2017) and projections from twelve downscaled regional climate models to estimate future illness due to cryptosporidiosis and giardiasis from 2020-2099 in Vancouver, British Columbia, Canada.

APPENDIX: ADDITIONAL TABLES FOR CHAPTER 3

TABLE 25. AGI outcome case definitions (refer to TABLE 26 for complete list of ICD-9-CM diagnoses codes and definitions).

Case definition	ICD-9-CM codes	Descriptions
All-cause AGI	001.xx - 009.xx	GI illness
	558.9	Non-infectious GI illness
	787.91	Diarrhea NOS or nausea, vomiting, and diarrhea
	787.01	Nausea with vomiting
	787.02	Nausea alone
	787.03	Vomiting alone
All-cause AGI excluding *C. difficile* (*C. diff*) and cholera	Same as All-cause AGI above, but excludes: 001.xx, 008.45	*Excludes:* 001.xx cholera 008.45 *C. difficile*
All-cause AGI excluding nausea and/or vomiting (787.01-.03)	001.xx - 009.xx	GI illness
	558.9	Non-infectious GI illness
	787.91	Diarrhea NOS or nausea, vomiting, and diarrhea
Infectious AGI (excludes 007.3)	001.xx - 009.xx except 007.3*	GI illness *except trichomonas vaginalis – sexually-transmitted disease
Bacterial AGI	001.xx	cholera
	002.xx	Typhoid and paratyphoid fevers
	003.xx	Other salmonella infections. *Includes:* infection or food poisoning by Salmonella [any serotype]
	004.xx	Shigellosis (Includes: bacillary dysentery)
	005.xx	Other food poisoning (bacterial). *Excludes:* salmonella infections (003.0-003.9); toxic effect of: food contaminants (989.7), noxious foodstuffs (988.0-988.9)
	008.0x	Escherichia coli [E. coli]
	008.1	Arizona group of paracolon bacilli
	008.2	Aerobacter aerogenes; Enterobacter aerogenes
	008.3	Proteus (mirabilis) (morganii)
	008.4x	Other specified bacteria (Includes: Staphylococcus, Pseudomonas, Campylobacter, Yersinia enterocolitica, C. difficile (008.45), Other anaerobes, Other gram negative-bacteria, Other)
	008.5	Bacterial enteritis, unspecified
Protozoal AGI	006.xx	Amebiasis *Includes:* infection due to Entamoeba histolytica *Excludes:* amebiasis due to organisms other than Entamoeba histolytica (007.8)
	007.xx	Other protozoal intestinal diseases. *Includes:* protozoal: colitis, diarrhea, dysentery
Viral AGI	008.6x	Enteritis due to specified virus (Includes: Rotavirus, Adenovirus, Norovirus, Other small round viruses [SRV's], Calicivirus, Astrovirus, Enterovirus NEC, Other viral enteritis, Torovirus)
Other case definitions to consider in sensitivity analysis:		
Rotavirus	008.61	Rotavirus
Norovirus	008.63	Norwalk virus; Norovirus; Norwalk-like agent
C. difficile	008.45	Clostridium difficile; Pseudomembranous colitis
Other organism; Viral: enteritis NOS	008.8	Other organism, not elsewhere classified; Viral: enteritis NOS, gastroenteritis *Excludes:* influenza with involvement of gastrointestinal tract (487.8, 488.09, 488.19)
Other and unspecified noninfectious gastroenteritis and colitis	558.9	Other and unspecified noninfectious gastroenteritis and colitis
Diarrhea NOS	787.91	Diarrhea NOS

TABLE 26. Comparison of acute gastrointestinal illness (AGI) definitions using ICD-9-CM diagnosis codes. The case definitions for ICD-9-CM primary and secondary diagnosis codes for case definition for acute gastrointestinal illness (AGI) in proposed study (001-009, 558.9, 787.91, 787.01-.03). Note that this case definition expands the definition from Hartley (2016) by including ICD-9-CM codes for nausea with vomiting (787.01-787.03) (DeFelice, 2014; DeFelice et al., 2015; Tinker et al., 2009, 2010), but excludes other codes in Wade *et al.* (2014) **abdominal pain (789.0).**

ICD-9-CM diagnosis codes related to gastrointestinal illness(es)			Recent articles that use ICD-9-CM codes for AGI definitions First Author (year) in order by year of publication											
Code	**Sub-code**	**ICD-9-CM Diagnosis**	**Proposed AGI Definition**	**Gleason et al. (2017)**	**Hartley (2016)**	**DeFelice et al. (2014, 2015)**	**Wade et al. (2014)**	**Uejio et al. (2014)**	**Desai et al. (2012)**	**Lopman et al. (2011)**	**Tinker et al. (2009, 2010)**	**Redman et al. (2007)**	**Schwartz et al. (1997)**	**Gangarosa et al. (1992)**
001		Cholera	1	1	1	1	1	1	1	1	1	1	1	1
	001.0	Due to Vibrio cholera	1	1	1	1	1	1	1	1	1	1	1	1
	001.1	Due to Vibrio cholerae el tor	1	1	1	1	1	1	1	1	1	1	1	1
	001.9	Cholera, unspecified	1	1	1	1	1	1	1	1	1	1	1	1
002		Typhoid and paratyphoid fevers	1	1	1	1	1	1	1	1	1	1	1	1
	002.0	Typhoid fever; Typhoid (fever) (infection) [any site]	1	1	1	1	1	1	1	1	1	1	1	1
	002.1	Paratyphoid fever A	1	1	1	1	1	1	1	1	1	1	1	1
	002.2	Paratyphoid fever B	1	1	1	1	1	1	1	1	1	1	1	1
	002.3	Paratyphoid fever C	1	1	1	1	1	1	1	1	1	1	1	1
	002.9	Paratyphoid fever, unspecified	1	1	1	1	1	1	1	1	1	1	1	1
003		Other salmonella infections *Includes:* infection or food poisoning by Salmonella [any serotype]	1	1	1	1	1	$\frac{10}{11}$	$\frac{10}{11}$	1	1	1	1	1
	003.0	Salmonella gastroenteritis (Salmonellosis)	1	1	1	1	1	1	1	1	1	1	1	1
	003.1	Salmonella septicemia	1	1	1	1	1	1	1	1	1	1	1	1
	003.2	Localized salmonella infections	1	1	1	1	1	1	1	1	1	1	1	1
	003.20	Localized salmonella infection, unspecified	1	1	1	1	1	0	0	1	1	1	1	1
	003.21	Salmonella meningitis	1	1	1	1	1	1	1	1	1	1	1	1
	003.22	Salmonella pneumonia	1	1	1	1	1	1	1	1	1	1	1	1
	003.23	Salmonella arthritis	1	1	1	1	1	1	1	1	1	1	1	1

ICD-9-CM diagnosis codes related to gastrointestinal illness(es)			Recent articles that use ICD-9-CM codes for AGI definitions First Author (year) in order by year of publication											
Code	**Sub-code**	**ICD-9-CM Diagnosis**	**Proposed AGI Definition**	**Gleason et al. (2017)**	**Hartley (2016)**	**DeFelice et al. (2014, 2015)**	**Wade et al. (2014)**	**Uejio et al. (2014)**	**Desai et al. (2012)**	**Lopman et al. (2011)**	**Tinker et al. (2009, 2010)**	**Redman et al. (2007)**	**Schwartz et al. (1997)**	**Gangarosa et al. (1992)**
	003.24	Salmonella osteomyelitis	1	1	1	1	1	1	1	1	1	1	1	1
	003.29	Other	1	1	1	1	1	1	1	1	1	1	1	1
	003.8	Other specified salmonella infections	1	1	1	1	1	1	1	1	1	1	1	1
	003.9	Salmonella infection, unspecified	1	1	1	1	1	1	1	1	1	1	1	1
004		Shigellosis (Includes: bacillary dysentery)	1	1	1	1	1	1	1	1	1	1	1	1
	004.0	Shigella dysenteriae [Infection by group A Shigella (Schmitz) (Shiga)]	1	1	1	1	1	1	1	1	1	1	1	1
	004.1	Shigella flexneri [Infection by group B Shigella]	1	1	1	1	1	1	1	1	1	1	1	1
	004.2	Shigella boydii [Infection by group C Shigella]	1	1	1	1	1	1	1	1	1	1	1	1
	004.3	Shigella sonnei [Infection by group D Shigella]	1	1	1	1	1	1	1	1	1	1	1	1
	004.8	Other specified shigella infections	1	1	1	1	1	1	1	1	1	1	1	1
	004.9	Shigellosis, unspecified	1	1	1	1	1	1	1	1	1	1	1	1
005		Other food poisoning (bacterial) *Excludes:* salmonella infections (003.0-003.9); toxic effect of: food contaminants (989.7), noxious foodstuffs (988.0-988.9)	1	1	1	1	1	$\frac{2}{9}$	1	1	$\frac{4}{9}$	1	1	1
	005.0	Staphylococcal food poisoning; Staphylococcal toxemia specified as due to food	1	1	1	1	1	0	1	1	1	1	1	1
	005.1	Botulism food poisoning; Botulism NOS; Food poisoning due to Clostridium botulinum *Excludes:* infant botulism (040.41), wound botulism (040.42)	1	1	1	1	1	0	1	1	0	1	1	1
	005.2	Food poisoning due to Clostridium perfringens [C. welchii]; Enteritis necroticans	1	1	1	1	1	0	1	1	0	1	1	1
	005.3	Food poisoning due to other Clostridia	1	1	1	1	1	0	1	1	0	1	1	1
	005.4	Food poisoning due to Vibrio parahaemolyticus	1	1	1	1	1	1	1	1	1	1	1	1
	005.8	Other bacterial food poisoning *Excludes:* salmonella food poisoning (003.0-003.9)	1	1	1	1	1	0	1	1	0	1	1	1
	005.81	Food poisoning due to Vibrio vulnificus	1	1	1	1	1	1	1	1	0	1	1	1

ICD-9-CM diagnosis codes related to gastrointestinal illness(es)			Recent articles that use ICD-9-CM codes for AGI definitions First Author (year) in order by year of publication											
Code	**Sub-code**	**ICD-9-CM Diagnosis**	**Proposed AGI Definition**	Gleason et al. (2017)	Hartley (2016)	DeFelice et al. (2014, 2015)	Wade et al. (2014)	Uejio et al. (2014)	Desai et al. (2012)	Lopman et al. (2011)	Tinker et al. (2009, 2010)	Redman et al. (2007)	Schwartz et al. (1997)	Gangarosa et al. (1992)
	005.89	Other bacterial food poisoning; Food poisoning due to Bacillus cereus	1	1	1	1	1	0	1	1	1	1	1	1
	005.9	Food poisoning, unspecified	1	1	1	1	1	0	1	1	1	1	1	1
006		Amebiasis *Includes:* infection due to Entamoeba histolytica *Excludes:* amebiasis due to organisms other than Entamoeba histolytica (007.8)	1	1	1	1	1	1	$\frac{5}{11}$	$\frac{4}{11}$	1	1	1	$\frac{1}{11}$
	006.0	Acute amebic dysentery without mention of abscess; Acute amebiasis	1	1	1	1	1	1	1	1	1	1	1	1
	006.1	Chronic intestinal amebiasis without mention of abscess; Chronic: amebiasis, amebic dysentery	1	1	1	1	1	1	1	1	1	1	1	0
	006.2	Amebic nondysenteric colitis	1	1	1	1	1	1	1	1	1	1	1	0
	006.3	Amebic liver abscess; Hepatic amebiasis	1	1	1	1	1	1	0	0	1	1	1	0
	006.4	Amebic lung abscess; Amebic abscess of lung (and liver)	1	1	1	1	1	1	0	0	1	1	1	0
	006.5	Amebic brain abscess; Amebic abscess of brain (and liver) (and lung)	1	1	1	1	1	1	0	0	1	1	1	0
	006.6	Amebic skin ulceration; Cutaneous amebiasis	1	1	1	1	1	1	0	0	1	1	1	0
	006.8	Amebic infection of other sites; Amebic: appendicitis, balanitis; Ameboma *Excludes:* specific infections by free-living amebae (136.21-136-29)	1	1	1	1	1	1	1	0	1	1	1	0
	006.9	Amebiasis, unspecified; Amebiasis NOS	1	1	1	1	1	1	1	1	1	1	1	0
007		Other protozoal intestinal diseases *Includes:* protozoal: colitis, diarrhea, dysentery	1	1	1	1	1	1	1	1	1	1	1	1
	007.0	Balantidiasis; Infection by Balantidium coli	1	1	1	1	1	1	1	1	1	1	1	1
	007.1	Giardiasis; Infection by Giardia lamblia; Lambliasis	1	1	1	1	1	1	1	1	1	1	1	1
	007.2	Coccidiosis; Infection by Isospora belli and Isospora hominis; Isosporiasis	1	1	1	1	1	1	1	1	1	1	1	1
	007.3	Intestinal trichomoniasis	1	1	1	1	1	1	1	1	1	1	1	1
	007.4	Cryptosporidiosis	1	1	1	1	1	1	1	1	1	1	1	1
	007.5	Cyclosporiasis	1	1	1	1	1	1	1	1	1	1	1	1

ICD-9-CM diagnosis codes related to gastrointestinal illness(es)			Recent articles that use ICD-9-CM codes for AGI definitions First Author (year) in order by year of publication											
Code	**Sub-code**	**ICD-9-CM Diagnosis**	**Proposed AGI Definition**	Gleason et al. (2017)	Hartley (2016)	DeFelice et al. (2014, 2015)	Wade et al. (2014)	Uejio et al. (2014)	Desai et al. (2012)	Lopman et al. (2011)	Tinker et al. (2009, 2010)	Redman et al. (2007)	Schwartz et al. (1997)	Gangarosa et al. (1992)
	007.8	Other specified protozoal intestinal diseases; Amebiasis due to organisms other than Entamoeba histolytica	1	1	1	1	1	1	1	1	1	1	1	1
	007.9	Unspecified protozoal intestinal disease; Flagellate diarrhea; Protozoal dysentery NOS	1	1	1	1	1	1	1	1	1	1	1	1
008		Intestinal infections due to other organisms *Includes:* any condition classifiable to 009.0-009.3 with mention of the responsible organisms *Excludes:* food poisoning by these organisms (005.0-005.9)	1	1	1	1	29/30	1	1	29/30	23/30	1	1	1
	008.0	Escherichia coli [E. coli]	1	1	1	1	1	1	1	1	1	1	1	1
	008.00	E. coli, unspecified; E. coli enteritis NOS	1	1	1	1	1	1	1	1	1	1	1	1
	008.01	Enteropathogenic E. coli	1	1	1	1	1	1	1	1	1	1	1	1
	008.02	Enterotoxigenic E. coli	1	1	1	1	1	1	1	1	1	1	1	1
	008.03	Enteroinvasive E. coli	1	1	1	1	1	1	1	1	1	1	1	1
	008.04	Enterohemorrhagic E. coli	1	1	1	1	1	1	1	1	1	1	1	1
	008.09	Other intestinal E. coli infections	1	1	1	1	1	1	1	1	1	1	1	1
	008.1	Arizona group of paracolon bacilli	1	1	1	1	1	1	1	1	0	1	1	1
	008.2	Aerobacter aerogenes; Enterobacter aerogenes	1	1	1	1	1	1	1	1	0	1	1	1
	008.3	Proteus (mirabilis) (morganii)	1	1	1	1	1	1	1	1	0	1	1	1
	008.4	Other specified bacteria	1	1	1	1	1	1	1	1	0	1	1	1
	008.41	Staphylococcus; Staphylococcal enterocolitis	1	1	1	1	1	1	1	1	0	1	1	1
	008.42	Pseudomonas	1	1	1	1	1	1	1	1	1	1	1	1
	008.43	Campylobacter	1	1	1	1	1	1	1	1	1	1	1	1
	008.44	Yersinia enterocolitica	1	1	1	1	1	1	1	1	1	1	1	1
	008.45	Clostridium difficile; Pseudomembranous colitis	1	1	1	1	0	1	1	0	0	1	1	1

ICD-9-CM diagnosis codes related to gastrointestinal illness(es)			Recent articles that use ICD-9-CM codes for AGI definitions First Author (year) in order by year of publication											
Code	**Sub-code**	**ICD-9-CM Diagnosis**	**Proposed AGI Definition**	**Gleason et al. (2017)**	**Hartley (2016)**	**DeFelice et al. (2014, 2015)**	**Wade et al. (2014)**	**Uejio et al. (2014)**	**Desai et al. (2012)**	**Lopman et al. (2011)**	**Tinker et al. (2009, 2010)**	**Redman et al. (2007)**	**Schwartz et al. (1997)**	**Gangarosa et al. (1992)**
	008.46	Other anaerobes; Anaerobic enteritis NOS; Bacteroides (fragilis); Gram-negative anaerobes	1	1	1	1	1	1	1	1	0	1	1	1
	008.47	Other gram-negative bacteria; Gram-negative enteritis NOS *Excludes:* gram-negative anaerobes (008.46)	1	1	1	1	1	1	1	1	1	1	1	1
	008.49	Other	1	1	1	1	1	1	1	1	1	1	1	1
	008.5	Bacterial enteritis, unspecified	1	1	1	1	1	1	1	1	1	1	1	1
	008.6	Enteritis due to specified virus	1	1	1	1	1	1	1	1	1	1	1	1
	008.61	Rotavirus	1	1	1	1	1	1	1	1	1	1	1	1
	008.62	Adenovirus	1	1	1	1	1	1	1	1	1	1	1	1
	008.63	Norwalk virus; Norovirus; Norwalk-like agent	1	1	1	1	1	1	1	1	1	1	1	1
	008.64	Other small round viruses [SRV's]; Small round virus NOS	1	1	1	1	1	1	1	1	1	1	1	1
	008.65	Calicivirus	1	1	1	1	1	1	1	1	1	1	1	1
	008.66	Astrovirus	1	1	1	1	1	1	1	1	1	1	1	1
	008.67	Enterovirus NEC; Coxsackie virus; Echovirus *Excludes:* poliovirus (045.0-045.9)	1	1	1	1	1	1	1	1	1	1	1	1
	008.69	Other viral enteritis; Torovirus	1	1	1	1	1	1	1	1	1	1	1	1
	008.8	Other organism, not elsewhere classified; Viral: enteritis NOS, gastroenteritis *Excludes:* influenza with involvement of gastrointestinal tract (487.8, 488.09, 488.19)	1	1	1	1	1	1	1	1	1	1	1	1

ICD-9-CM diagnosis codes related to gastrointestinal illness(es)			Recent articles that use ICD-9-CM codes for AGI definitions First Author (year) in order by year of publication											
Code	**Sub-code**	**ICD-9-CM Diagnosis**	**Proposed AGI Definition**	**Gleason et al. (2017)**	**Hartley (2016)**	**DeFelice et al. (2014, 2015)**	**Wade et al. (2014)**	**Uejio et al. (2014)**	**Desai et al. (2012)**	**Lopman et al. (2011)**	**Tinker et al. (2009, 2010)**	**Redman et al. (2007)**	**Schwartz et al. (1997)**	**Gangarosa et al. (1992)**
009		Ill-defined intestinal infections *Excludes:* diarrheal disease or intestinal infection due to specified organism (001.0-008.8); diarrhea following gastrointestinal surgery (564.4); intestinal malabsorption (579.0-579.9); ischemic enteritis (557.0-557.9); other noninfectious gastroenteritis and colitis (558.1-558.9); regional enteritis (555.0-555.9); ulcerative colitis (556)	1	1	1	1	1	1	1	1	1	1	1	1
	009.0	Infectious colitis, enteritis, and gastroenteritis; Colitis (septic); Dysentery: NOS, catarrhal, hemorrhagic; Enteritis (septic); Gastroenteritis (septic)	1	1	1	1	1	1	1	1	1	1	1	1
	009.1	Colitis, enteritis, and gastroenteritis of presumed infectious origin *Excludes:* colitis NOS (558.9); enteritis NOS (558.9); gastroenteritis NOS (558.9)	1	1	1	1	1	1	1	1	1	1	1	1
	009.2	Infectious diarrhea; Diarrhea: dysenteric, epidemic; Infectious diarrheal disease NOS	1	1	1	1	1	1	1	1	1	1	1	1
	009.3	Diarrhea of presumed infectious origin *Excludes:* diarrhea NOS (787.91)	1	1	1	1	1	1	1	1	1	1	1	1
	558.9	Other and unspecified noninfectious gastroenteritis and colitis	1	0	1	1	1	1	1	1	1	1	1	1
787		Symptoms involving digestive system *Excludes:* constipation (564.0-564.9); pylorospasm (537.81); congenital (750.5)	$\frac{4}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{4}{15}$	$\frac{8}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{4}{15}$	$\frac{1}{15}$	0	$\frac{1}{15}$
	787.0	Nausea and vomiting; Emesis *Excludes:* hematemesis NOS (578.0); vomiting: bilious, following gastrointestinal surgery (564.3), cyclical (536.2), associated with migraine (346.2), fecal matter (569.87), psychogenic (306.4), excessive, in pregnancy (643.0-643.9), habit (536.2), of newborn (779.32, 779.33), persistent (536.2), psychogenic NOS (307.54)	0	0	0	0	1	0	0	0	0	0	0	0
	787.01	Nausea with vomiting	1	0	0	1	1	0	0	0	1	0	0	0
	787.02	Nausea alone	1	0	0	1	1	0	0	0	1	0	0	0
	787.03	Vomiting alone*	1	0	0	1	1	0	0	0	1	0	0	1

ICD-9-CM diagnosis codes related to gastrointestinal illness(es)			Recent articles that use ICD-9-CM codes for AGI definitions First Author (year) in order by year of publication											
Code	**Sub-code**	**ICD-9-CM Diagnosis**	**Proposed AGI Definition**	**Gleason et al. (2017)**	**Hartley (2016)**	**DeFelice et al. (2014, 2015)**	**Wade et al. (2014)**	**Uejio et al. (2014)**	**Desai et al. (2012)**	**Lopman et al. (2011)**	**Tinker et al. (2009, 2010)**	**Redman et al. (2007)**	**Schwartz et al. (1997)**	**Gangarosa et al. (1992)**
	787.04	Bilious emesis; Bilious vomiting *Excludes:* bilious emesis (vomiting) in newborn (779.32)	0	0	0	0	1	0	0	0	0	0	0	0
	787.1	Heartburn; Pyrosis; Waterbrash *Excludes:* dyspepsia or indigestion (536.8)	0	0	0	0	0	0	0	0	0	0	0	0
	787.2	Dysphagia	0	0	0	0	0	0	0	0	0	0	0	0
	787.3	Flatulence, eructation, and gas pain; Abdominal distention (gaseous); Bloating; Tympanites (abdominal) (intestinal) *Excludes:* aerophagy (306.4)	0	0	0	0	0	0	0	0	0	0	0	0
	787.4	Visible peristalsis; Hyperperistalsis	0	0	0	0	1	0	0	0	0	0	0	0
	787.5	Abnormal bowel sounds; Absent bowel sounds; Hyperactive bowel sounds	0	0	0	0	0	0	0	0	0	0	0	0
	787.6	Incontinence of feces; Encopresis NOS; Incontinence of sphincter ani *Excludes:* that of nonorganic origin (307.7)	0	0	0	0	0	0	0	0	0	0	0	0
	787.7	Abnormal feces; Bulky stools *Excludes:* abnormal stool content (792.1); melena: NOS (578.1), newborn (772.4, 777.3)	0	0	0	0	0	0	0	0	0	0	0	0
	787.9	Other symptoms involving digestive system; *Excludes:* gastrointestinal hemorrhage (578.0-578.9); intestinal obstruction (560.0-560.9); specific functional digestive disorders: esophagus (530.0-530.9), stomach and duodenum (536.0-536.9), those not elsewhere classified (564.0-564.9),	0	0	0	0	1	0	0	0	0	0	0	0
	787.91	Diarrhea, Diarrhea NOS	1	1	1	1	1	1	1	1	1	1	0	0
	787.99	Other; Change in bowel habits; Tenesmus (rectal)	0	0	0	0	0	0	0	0	0	0	0	0

** Gangarosa et al. (1992) used 078.82 epidemic or winter vomiting, which converts to vomiting without nausea (787.03)*

Source: ICD-9-CM codes from 2011 revision from the DTAB11.ZIP file available at: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Publications/ICD9-CM/2010

DISCLAIMERS

The North Carolina Disease Event Tracking and Epidemiologic Collection Tool (NC DETECT) is an advanced, statewide public health surveillance system. NC DETECT is funded with federal funds by North Carolina Division of Public Health (NC DPH), Public Health Emergency Preparedness Grant (PHEP), and managed through a collaboration between NC DPH and the University of North Carolina at Chapel Hill Department of Emergency Medicine's Carolina Center for Health Informatics (UNC CCHI). The NC DETECT Data Oversight Committee does not take responsibility for the scientific validity or accuracy of methodology, results, statistical analyses, or conclusions presented.

REFERENCES

Adhikari, R., & Agrawal, R. (2013). An introductory study on time series modeling and forecasting. ArXiv:1302.6613, 67. https://doi.org/10.13140/2.1.2771.8084

Agresti, A. (2002). *Categorical Data Analysis* (2nd ed.). John Wiley & Sons, Inc. https://doi.org/10.1002/0471249688

Ahmed, S. M., Lopman, B. A., & Levy, K. (2013). A systematic review and meta-analysis of the global seasonality of norovirus. *PLoS ONE*, 8(10), e75922. https://doi.org/10.1371/journal.pone.0075922

Aik, J., Heywood, A. E., Newall, A. T., Ng, L. C., Kirk, M. D., & Turner, R. (2018). Climate variability and salmonellosis in Singapore – A time series analysis. *Science of the Total Environment*, 639, 1261–1267. https://doi.org/10.1016/j.scitotenv.2018.05.254

Aik, J., Ong, J., & Ng, L. C. (2020). The effects of climate variability and seasonal influence on diarrhoeal disease in the tropical city-state of Singapore – A time-series analysis. *International Journal of Hygiene and Environmental Health*, 227. https://doi.org/10.1016/j.ijheh.2020.113517

Alarcon Falconi, T. M., Estrella, B., Sempértegui, F., & Naumova, E. N. (2020). Effects of data aggregation on time series analysis of seasonal infections. *International Journal of Environmental Research and Public Health*, 17(16), 5887. https://doi.org/10.3390/ijerph17165887

Alexander, K. A., Heaney, A. K., & Shaman, J. (2018). Hydrometeorology and flood pulse dynamics drive diarrheal disease outbreaks and increase vulnerability to climate change in surface-water-dependent populations: A retrospective analysis. *PLOS Medicine*, 15(11), e1002688. https://doi.org/10.1371/journal.pmed.1002688

Ali, M., Kim, D. R., Yunus, M., & Emch, M. (2013). Time series analysis of cholera in Matlab, Bangladesh, during 1988-2001. *Journal of Health, Population and Nutrition*, 31(1), 11–19. https://doi.org/10.3329/jhpn.v31i1.14744

Allard, R., Plante, C., Garnier, C., & Kosatsky, T. (2011). The reported incidence of campylobacteriosis modelled as a function of earlier temperatures and numbers of cases, Montreal, Canada, 1990–2006. *International Journal of Biometeorology*, 55(3), 353–360. https://doi.org/10.1007/s00484-010-0345-x

Almon, S. (1965). The distributed lag between capital appropriations and expenditures. *Econometrica*, 33(1), 178–196. https://doi.org/10.2307/1911894

Altizer, S., Dobson, A., Hosseini, P., Hudson, P., Pascual, M., & Rohani, P. (2006). Seasonality and the dynamics of infectious diseases. *Ecology Letters*, 9(4), 467–484. https://doi.org/10.1111/j.1461-0248.2005.00879.x

Aminharati, F., Soltan Dallal, M. M., Ehrampoush, M. H., Dehghani-Tafti, A., Yaseri, M., Memariani, M., & Rajabi, Z. (2018). The effect of environmental parameters on the incidence of Shigella outbreaks in Yazd province, Iran. *Water Supply*, 18(4), 1388–1395. https://doi.org/10.2166/ws.2017.205

Anderson, A. D., Heryford, A. G., Sarisky, J. P., Higgins, C., Monroe, S. S., Beard, R. S., Newport, C. M., Cashdollar, J. L., Fout, G. S., Robbins, D. E., Seys, S. A., Musgrave, K. J., Medus, C., Vinjé, J., Bresee, J. S., Mainzer, H. M., & Glass, R. I. (2003). A waterborne outbreak of Norwalk-like virus among snowmobilers - Wyoming, 2001. *Journal of Infectious Diseases*, 187(2), 303–306. https://doi.org/10.1086/346239

Anthoff, D., & Tol, R. S. J. (2013a). The Climate Framework for Uncertainty, Negotiation and Distribution (Fund), Tables, Version 3.7. In www.fund-model.org. http://www.fund-model.org/versions

Anthoff, D., & Tol, R. S. J. (2013b). The Climate Framework for Uncertainty, Negotiation and Distribution (Fund), Technical Description, Version 3.7. In www.fund-model.org. http://www.fund-model.org/versions

Armstrong, B. (2006). Models for the relationship between ambient temperature and daily mortality. *Epidemiology*, 17(6), 624–631. https://doi.org/10.1097/01.ede.0000239732.50999.8f

Atchison, C. J., Tam, C. C., Hajat, S., van Pelt, W., Cowden, J. M., & Lopman, B. A. (2010). Temperature-dependent transmission of rotavirus in Great Britain and the Netherlands. *Proceedings of the Royal Society B: Biological Sciences*, 277(1683), 933–942. https://doi.org/10.1098/rspb.2009.1755

Auffhammer, M., Hsiang, S. M., Schlenker, W., & Sobel, A. (2013). Using Weather Data and Climate Model Output in Economic Analyses of Climate Change. *Review of Environmental Economics and Policy*, 7(2), 181–198. http://reep.oxfordjournals.org/content/7/2/181.abstract

Azage, M., Kumie, A., Worku, A., Bagtzoglou, A. C., & Anagnostou, E. (2017). Effect of climatic variability on childhood diarrhea and its high risk periods in northwestern parts of Ethiopia. *PLOS ONE*, 12(10), e0186933. https://doi.org/10.1371/JOURNAL.PONE.0186933

Baccini, M., Biggeri, A., Lagazio, C., Lertxundi, A., & Saez, M. (2007). Parametric and semi-parametric approaches in the analysis of short-term effects of air pollution on health. *Computational Statistics and Data Analysis*, 51(9), 4324–4336. https://doi.org/10.1016/j.csda.2006.05.026

Bach, P. M., McCarthy, D. T., & Deletic, A. (2010). Redefining the stormwater first flush phenomenon. *Water Research*, 44(8), 2487–2498. https://doi.org/10.1016/j.watres.2010.01.022

Bandyopadhyay, S., Kanji, S., & Wang, L. (2012). The impact of rainfall and temperature variation on diarrheal prevalence in Sub-Saharan Africa. *Applied Geography*, 33, 63–72. https://doi.org/10.1016/j.apgeog.2011.07.017

Barrett, J., & Fhogartaigh, C. N. (2017). Bacterial gastroenteritis. *Medicine*, 45(11), 683–689. https://doi.org/10.1016/j.mpmed.2017.08.002

Bateson, T. F., & Schwartz, J. (1999). Control for Seasonal Variation and Time Trend in Case-Crossover Studies of Acute Effects of Environmental Exposures. *Epidemiology*, 10(5), 539–544. https://doi.org/10.1097/00001648-199909000-00013

Beaudeau, P., Le Tertre, A., Zeghnoun, A., Zanobetti, A., & Schwartz, J. (2012). A time series study of drug sales and turbidity of tap water in Le Havre, France. *Journal of Water and Health*, 10(2), 221–235. https://doi.org/10.2166/wh.2012.157

Beaudeau, P., Schwartz, J., & Levin, R. (2014). Drinking water quality and hospital admissions of elderly people for gastrointestinal illness in Eastern Massachusetts, 1998-2008. *Water Research*, 52, 188–198. https://doi.org/10.1016/j.watres.2014.01.005

Beaudeau, P., Zeghnoun, A., Corso, M., Lefranc, A., & Rambaud, L. (2014). A time series study of gastroenteritis and tap water quality in the Nantes area, France, 2002-2007. *In Journal of Exposure Science and Environmental Epidemiology* (Vol. 24, Issue 2, pp. 192–199). Nature Publishing Group. https://doi.org/10.1038/jes.2013.5

Beck-Johnson, L. M., Nelson, W. A., Paaijmans, K. P., Read, A. F., Thomas, M. B., & Bjornstad, O. N. (2013). The Effect of Temperature on Anopheles Mosquito Population Dynamics and the Potential for Malaria Transmission. *Plos ONE*, 8(11), e79276–e79276. https://doi.org/10.1371/journal.pone.0079276

Bennett, D. G., & Patton, J. C. (2008). A Geography of the Carolinas. In *A Geography of the Carolinas*. Parkway Publishers, Inc. https://epdf.pub/a-geography-of-the-carolinas.html

Bertrand-Krajewski, J. L., Chebbo, G., & Saget, A. (1998). Distribution of pollutant mass vs volume in stormwater discharges and the first flush phenomenon. *Water Research*, 32(8), 2341–2356. https://doi.org/10.1016/S0043-1354(97)00420-X

Bhaskaran, K., Gasparrini, A., Hajat, S., Smeeth, L., & Armstrong, B. (2013). Time series regression studies in environmental epidemiology. *International Journal of Epidemiology*, 42(4), 1187–1195. https://doi.org/10.1093/ije/dyt092

Bhavnani, D., Goldstick, J. E., Cevallos, W., Trueba, G., & Eisenberg, J. N. S. (2014). Impact of rainfall on diarrheal disease risk associated with unimproved water and sanitation. *The American Journal of Tropical Medicine and Hygiene*, 90(4), 705–711. https://doi.org/10.4269/ajtmh.13-0371

Bi, P., Cameron, A. S., Zhang, Y., & Parton, K. A. (2008). Weather and notified Campylobacter infections in temperate and sub-tropical regions of Australia: An ecological study. *Journal of Infection*, 57(4), 317–323. https://doi.org/10.1016/j.jinf.2008.08.004

Bifolchi, N., Michel, P., Talbot, J., Svenson, L., Simmonds, K., Checkley, S., Chui, L., Dick, P., & Wilson, J. B. (2014). Weather and livestock risk factors for Escherichia coli O157 human infection in Alberta, Canada. *Epidemiology and Infection*, 142(11), 2302–2313. https://doi.org/10.1017/S0950268813002781

Boithias, L., Choisy, M., Souliyaseng, N., Jourdren, M., Quet, F., Buisson, Y., Thammahacksa, C., Silvera, N., Latsachack, K., Sengtaheuanghoung, O., Pierret, A., Rochelle-Newall, E., Becerra, S., & Ribolzi, O. (2016). Hydrological regime and water shortage as drivers of the seasonal incidence of diarrheal diseases in a tropical montane environment. *PLOS Neglected Tropical Diseases*, 10(12), e0005195. https://doi.org/10.1371/journal.pntd.0005195

Borchardt, M. A., Chyou, P.-H., DeVries, E. O., & Belongia, E. A. (2003). Septic system density and infectious diarrhea in a defined population of children. *Environmental Health Perspectives*, 111(5), 742–748. https://doi.org/10.1289/ehp.5914

Bouma, M. J., & Pascual, M. (2001). Seasonal and interannual cycles of endemic cholera in Bengal 1891–1940 in relation to climate and geography. *Hydrobiologia*, 460(1–3), 147–156. https://doi.org/10.1023/A:1013165215074

Box, G. E. P., & Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control* (1st ed.). Holden Day.

Brabec, E., Schulte, S., & Richards, P. L. (2002). Impervious surfaces and water quality: A review of current literature and its implications for watershed planning. *Journal of Planning Literature*, 16(4), 499–514. https://doi.org/10.1177/088541202400903563

Bradatan, C., Dennis, J. A., Flores-Yeffal, N., & Swain, S. (2020). Child health, household environment, temperature and rainfall anomalies in Honduras: A socio-climate data linked analysis. *Environmental Health: A Global Access Science Source*, 19(1), 1–12. https://doi.org/10.1186/s12940-020-0560-9

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and Regression Trees*. Wadsworth & Brooks.

Briant, A., Combes, P.-P., & Lafourcade, M. (2010). Dots to boxes: Do the size and shape of spatial units jeopardize economic geography estimations? *Journal of Urban Economics*, 67(3), 287–302. https://doi.org/10.1016/j.jue.2009.09.014

Britton, E., Hales, S., Venugopal, K., & Baker, M. G. (2010). Positive association between ambient temperature and salmonellosis notifications in New Zealand, 1965-2006. *Australian and New Zealand Journal of Public Health*, 34(2), 126–129. https://doi.org/10.1111/j.1753-6405.2010.00495.x

Brockwell, P. J., & Davis, R. A. (2002). *Introduction to Time Series and Forecasting*. Springer.

Brown, C., Meeks, R., Ghile, Y., & Hunu, K. (2013). Is water security necessary ? An empirical analysis of the effects of climate hazards on national-level economic growth. *Phil Trans R Soc A*, 371(20120416), 1–18. https://doi.org/10.1098/rsta.2012.0416

Brubacher, J., Allen, D. M., Déry, S. J., Parkes, M. W., Chhetri, B., Mak, S., Sobie, S., & Takaro, T. K. (2020). Associations of five food- and water-borne diseases with ecological zone, land use and aquifer type in a changing climate. *Science of the Total Environment*, 728, 138808. https://doi.org/10.1016/j.scitotenv.2020.138808

Bunch, M., Morrison, K., Parkes, M., & Venema, H. (2011). Promoting Health and Well-Being by Managing for Social–Ecological Resilience: the Potential of Integrating Ecohealth and Water Resources Management Approaches. *Ecology and Society*, 16(1), art6.

Burkholder, J. A., Libra, B., Weyer, P., Heathcote, S., Kolpin, D., Thorne, P. S., & Wichman, M. (2007). Impacts of waste from concentrated animal feeding operations on water quality. *Environmental Health Perspectives*, 115(2), 308–312. https://doi.org/10.1289/ehp.8839

Busch, J., Berrang-Ford, L., Clark, S., Patterson, K., Windfeld, E., Donnelly, B., Lwasa, S., Namanya, D., & Harper, S. L. (2019). Is the effect of precipitation on acute gastrointestinal illness in southwestern Uganda different between Indigenous and non-Indigenous communities? *PLOS ONE*, 14(5), e0214116. https://doi.org/10.1371/journal.pone.0214116

Bush, K. F., O'Neill, M. S., Li, S., Mukherjee, B., Hu, H., Ghosh, S., & Balakrishnan, K. (2014b). Associations between extreme precipitation and gastrointestinal-related hospital admissions in Chennai, India. *Environmental Health Perspectives*, 122(3), 249–254. https://doi.org/10.1289/ehp.1306807

Bush, K., Fossani, C., Li, S., Mukherjee, B., Gronlund, C., & O'Neill, M. (2014a). Extreme precipitation and beach closures in the Great Lakes Region: Evaluating risk among the elderly. *International Journal of Environmental Research and Public Health*, 11(2), 2014–2032. https://doi.org/10.3390/ijerph110202014

Butler, A. J. (2013). Salmonella , weather and climate change in Australia [PhD Dissertation]. The Australian National University, Canberra, Australia.

Butler, A. J., & Hall, G. (2009). Methodological issues in the study of climate and gastroenteritis. IOP Conference Series: Earth and Environmental Science, 6(14), 142016.

Bylund, J., Toljander, J., Lysén, M., Rasti, N., Engqvist, J., & Simonsson, M. (2017). Measuring sporadic gastrointestinal illness associated with drinking water – an overview of methodologies. *Journal of Water and Health*, 15(3), 321–340. https://doi.org/10.2166/wh.2017.261

Cambrea, S. C., Petcu, L. C., Mihai, C. M., Hangan, T. L., & Iliescu, D. M. (2019). Influence of environmental factors about evolution of shigellosis in Constanta county of Romania. *Journal of Environmental Protection and Ecology*, 20(2), 986–994. https://docs.google.com/a/jepejournal.info/viewer?a=v&pid=sites&srcid=amVwZS1qb3VybmFsLmluZm98amVwZS1qb3VybmFsfGd4OjMwNmVmN2QzMGM5YTRjNzI

Cameron, A. C., & Trivedi, P. K. (1998). *Regression Analysis of Count Data (1st ed.)*. Cambridge University Press. www.cup.org/titles/catalogue.asp?isbn=0521635675

Cameron, A. C., & Trivedi, P. K. (2013). *Regression Analysis of Count Data (2nd ed.)*. Cambridge University Press.

Carey, V. (2019). gee: Generalized Estimation Equation Solver. R package version 4.13-25. https://cran.r-project.org/package=gee

Carlton, E. J., Eisenberg, J. N. S., Goldstick, J., Cevallos, W., Trostle, J., & Levy, K. (2014). Heavy rainfall events and diarrhea incidence: The role of social and environmental factors. *American Journal of Epidemiology,* 179(3), 344–352. https://doi.org/10.1093/aje/kwt279

Carlton, E. J., Woster, A. P., DeWitt, P., Goldstein, R. R. S., & Levy, K. (2016). A systematic review and meta-analysis of ambient temperature and diarrhoeal diseases. *International Journal of Epidemiology*, 45(1), 117–130. https://doi.org/10.1093/ije/dyv296

Carolina Center for Health Informatics, University of North Carolina at Chapel Hill. (2010). The UNC Department of Emergency Medicine Carolina Center for Health Informatics Report, NC DETECT Emergency Department Data: 2008. http://www.ncdetect.org/pubs.html

Carracedo-Martínez, E., Taracido, M., Tobias, A., Saez, M., & Figueiras, A. (2010). Case-crossover analysis of air pollution health effects: A systematic review of methodology and application. *Environmental Health Perspectives*, 118(8), 1173–1182. https://doi.org/10.1289/ehp.0901485

Carroll, B., Balogh, R., Morbey, H., & Araoz, G. (2010). Health and social impacts of a flood disaster: responding to needs and implications for practice. *Disasters*, 34(4), 1045–1063. https://doi.org/10.1111/j.1467-7717.2010.01182.x

Celik, C., Gozel, M. G., Turkay, H., Bakici, M. Z., Güven, A. S., & Elaldi, N. (2015). Rotavirus and adenovirus gastroenteritis: Time series analysis. *Pediatrics International*, 57(4), 590–596. https://doi.org/10.1111/ped.12592

Chadsuthi, S., Modchang, C., Lenbury, Y., Iamsirithaworn, S., & Triampo, W. (2012). Modeling seasonal leptospirosis transmission and its association with rainfall and temperature in Thailand using time–series and ARIMAX analyses. *Asian Pacific Journal of Tropical Medicine*, 5(7), 539–546. https://doi.org/10.1016/S1995-7645(12)60095-9

Chambers, J. M., & Hastie, T. (Eds.). (1991). *Statistical models in S*. Wadsworth & Brooks/Cole.

Chao, D. L., Roose, A., Roh, M., Kotloff, K. L., & Proctor, J. L. (2019). The seasonality of diarrheal pathogens: A retrospective study of seven sites over three years. *PLOS Neglected Tropical Diseases*, 13(8), e0007211. https://doi.org/10.1371/journal.pntd.0007211

Chatfield, C. (2001). *Time-Series Forecasting*. Chapman and Hall/CRC.

Checkley, W., Epstein, L. D., Gilman, R. H., Figueroa, D., Cama, R. I., Patz, J. A., & Black, R. E. (2000). Effects of El Nino and ambient temperature on hospital admissions for diarrhoeal diseases in Peruvian children. *The Lancet*, 355(9202), 442–450. https://doi.org/10.1016/S0140-6736(00)82010-3

Chen, C.-C., Lin, B.-C., Yap, L., Chiang, P.-H., & Chan, T.-C. (2018). The association between ambient temperature and acute diarrhea incidence in Hong Kong, Taiwan, and Japan. *Sustainability*, 10(5), 1417. https://doi.org/10.3390/su10051417

Chen, M.-J., Lin, C.-Y., Wu, Y.-T., Wu, P.-C., Lung, S.-C., & Su, H.-J. (2012). Effects of Extreme Precipitation to the Distribution of Infectious Diseases in Taiwan, 1994–2008. *PLoS ONE*, 7(6), e34651. https://doi.org/10.1371/journal.pone.0034651

Cheng, J., Xie, M. Y., Zhao, K. F., Wu, J. J., Xu, Z. W., Song, J., Zhao, D. S., Li, K. S., Wang, X., Yang, H. H., Wen, L. Y., Su, H., & Tong, S. L. (2017). Impacts of ambient temperature on the burden of bacillary dysentery in urban and rural Hefei, China. *Epidemiology and Infection*, 145(8), 1567–1576. https://doi.org/10.1017/S0950268817000280

Cheng, T., & Adepeju, M. (2014). Modifiable temporal unit problem (MTUP) and its effect on space-time cluster detection. *PLoS ONE*, 9(6), 1–10. https://doi.org/10.1371/journal.pone.0100465

Chhetri, B. K., Galanis, E., Sobie, S., Brubacher, J., Balshaw, R., Otterstatter, M., Mak, S., Lem, M., Lysyshyn, M., Murdock, T., Fleury, M., Zickfeld, K., Zubel, M., Clarkson, L., & Takaro, T. K. (2019). Projected local rain events due to climate change and the impacts on waterborne diseases in Vancouver, British Columbia, Canada. *Environmental Health*, 18(1), 116. https://doi.org/10.1186/s12940-019-0550-y

Chhetri, B. K., Takaro, T. K., Balshaw, R., Otterstatter, M., Mak, S., Lem, M., Zubel, M., Lysyshyn, M., Clarkson, L., Edwards, J., Fleury, M. D., Henderson, S. B., & Galanis, E. (2017a). Associations between extreme precipitation and acute gastro-intestinal illness due to cryptosporidiosis and giardiasis in an urban Canadian drinking water system (1997–2009). *Journal of Water and Health*, 15(6), 898–907. https://doi.org/10.2166/wh.2017.100

Chhetri, B. K., Takaro, T. K., Balshaw, R., Otterstatter, M., Mak, S., Lem, M., Zubel, M., Lysyshyn, M., Clarkson, L., Edwards, J., Fleury, M. D., Henderson, S. B., & Galanis, E. (2017b). Associations between extreme precipitation and acute gastro-intestinal illness due to cryptosporidiosis and giardiasis in an urban Canadian drinking water system (1997–2009). *Journal of Water and Health*, 15(6), 898–907. https://doi.org/10.2166/wh.2017.100

Chou, W.-C., Wu, J.-L., Wang, Y.-C., Huang, H., Sung, F.-C., & Chuang, C.-Y. (2010). Modeling the impact of climate variability on diarrhea-associated diseases in Taiwan (1996-2007). *Science of the Total Environment*, 409(1), 43–51. https://doi.org/10.1016/j.scitotenv.2010.09.001

Christenson, E. C., & Serre, M. L. (2017). Integrating remote sensing with nutrient management plans to calculate nitrogen parameters for swine CAFOs at the sprayfield and sub-watershed scales. *Science of the Total Environment*, 580, 865–872. https://doi.org/10.1016/j.scitotenv.2016.12.033

Chua, P. L. C., Ng, C. F. S., Tobias, A., Seposo, X. T., & Hashizume, M. (2021). Associations between ambient temperature and enteric infections by aetiology: A systematic review and meta-analysis [Preprint]. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3822570

Chua, P. L. C., Ng, C. F. S., Tobias, A., Seposo, X. T., & Hashizume, M. (2022). Associations between ambient temperature and enteric infections by pathogen: A systematic review and meta-analysis. *The Lancet Planetary Health*, 6(3), e202–e218. https://doi.org/10.1016/s2542-5196(22)00003-1

Chui, K. K. H., Jagai, J. S., Griffiths, J. K., & Naumova, E. N. (2011). Hospitalization of the elderly in the United States for nonspecific gastrointestinal diseases: A search for etiological clues. *American Journal of Public Health*, 101(11), 2082–2086. https://doi.org/10.2105/AJPH.2010.300096

Clark, L. A., & Pregibon, D. (1992). Tree-based models. In J. M. Chambers & T. J. Hastie (Eds.), *Statistical models in S* (pp. 377–419). Wadsworth.

Cline, M. (2020). Is North Carolina Rural or Urban? North Carolina Office of State Budget and Management (NC OSBM). https://www.osbm.nc.gov/blog/2020/11/19/north-carolina-rural-or-urban

Colford, J. M., Roy, S., Beach, M. J., Hightower, A., Shaw, S. E., & Wade, T. J. (2006). A review of household drinking water intervention trials and an approach to the estimation of endemic waterborne gastroenteritis in the United States. *Journal of Water and Health*, 4(S2), 71–88. https://doi.org/10.2166/wh.2006.018

Collier, S. A., Deng, L., Adam, E. A., Benedict, K. M., Beshearse, E. M., Blackstock, A. J., Bruce, B. B., Derado, G., Edens, C., Fullerton, K. E., Gargano, J. W., Geissler, A. L., Hall, A. J., Havelaar, A. H., Hill, V. R., Hoekstra, R. M., Reddy, S. C., Scallan, E., Stokes, E. K., … Beach, M. J. (2021). Estimate of burden and direct healthcare cost of infectious waterborne disease in the United States. *Emerging Infectious Diseases,* 27(1), 140–149. https://doi.org/10.3201/eid2701.190676

Colston, J. M., Ahmed, T., Mahopo, C., Kang, G., Kosek, M., de Sousa Junior, F., Shrestha, P. S., Svensen, E., Turab, A., & Zaitchik, B. (2018). Evaluating meteorological data from weather stations, and from satellites and global models for a multi-site epidemiological study. *Environmental Research*, 165(October 2017), 91–109. https://doi.org/10.1016/j.envres.2018.02.027

Colston, J. M., Zaitchik, B. F., Badr, H. S., Burnett, E., Ali, S. A., Rayamajhi, A., Satter, S. M., Eibach, D., Krumkamp, R., May, J., Chilengi, R., Howard, L. M., Sow, S. O., Jahangir Hossain, M., Saha, D., Imran Nisar, M., Zaidi, A. K. M., Kanungo, S., Mandomando, I., …

Kosek, M. N. (2022). Associations between eight earth observation-derived climate variables and enteropathogen infection: An independent participant data meta-analysis of surveillance studies with broad spectrum nucleic acid diagnostics. *GeoHealth*, 6(1), 1–20. https://doi.org/10.1029/2021GH000452

Colston, J. M., Zaitchik, B., Kang, G., Peñataro Yori, P., Ahmed, T., Lima, A., Turab, A., Mduma, E., Sunder Shrestha, P., Bessong, P., Peng, R. D., Black, R. E., Moulton, L. H., & Kosek, M. N. (2019). Use of earth observation-derived hydrometeorological variables to model and predict rotavirus infection (MAL-ED): a multisite cohort study. *The Lancet Planetary Health*, 3(6), e248–e258. https://doi.org/10.1016/S2542-5196(19)30084-1

Cook, S. M., Glass, R. I., LeBaron, C. W., & Ho, M.-S. (1990). Global seasonality of rotavirus infections. *Bulletin of the World Health Organization*, 68(2), 171–177. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2393128/

Corley, B., Bartelt-Hunt, S., Rogan, E., Coulter, D., Sparks, J., Baccaglini, L., Howell, M., Liaquat, S., Commack, R., & Kolok, A. S. (2018). Using watershed boundaries to map adverse health outcomes: Examples from Nebraska, USA. *Environmental Health Insights*, 12, 117863021775190. https://doi.org/10.1177/1178630217751906

Crane, S. R., & Moore, J. A. (1984). Bacterial pollution of groundwater: A review. *Water, Air, and Soil Pollution*, 22(1), 67–83. https://doi.org/10.1007/BF00587465

Craun, G. F., Brunkard, J. M., Yoder, J. S., Roberts, V. A., Carpenter, J., Wade, T., Calderon, R. L., Roberts, J. M., Beach, M. J., & Roy, S. L. (2010). Causes of outbreaks associated with drinking water in the United States from 1971 to 2006. *Clinical Microbiology Reviews*, 23(3), 507–528. https://doi.org/10.1128/CMR.00077-09

Croxen, M. A., Law, R. J., Scholz, R., Keeney, K. M., Wlodarska, M., & Finlay, B. B. (2013). Recent advances in understanding enteric pathogenic Escherichia coli. In Clinical Microbiology Reviews (Vol. 26, Issue 4, pp. 822–880). *American Society for Microbiology Journals*. https://doi.org/10.1128/CMR.00022-13

Cryer, J. D., & Chan, K.-S. (2008). *Time Series Analysis: With Applications to R*. https://doi.org/10.1007/978-0-387-75959-3

Cunliffe, N. A., Kilgore, P. E., Bresee, J. S., Steele, A. D., Luo, N., Hart, C. A., & Glass, R. I. (1998). Epidemiology of rotavirus diarrhoea in Africa: A review to assess the need for rotavirus immunization. *Bulletin of the World Health Organization*, 76(5), 525–537. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2305791/

Curriero, F. C., Patz, J. A., Rose, J. B., & Lele, S. (2001). The association between extreme precipitation and waterborne disease outbreaks in the United States, 1948-1994. *American Journal of Public Health*, 91(8), 1194–1199. https://doi.org/10.2105/AJPH.91.8.1194

Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology*, 28(15), 2031–2064. https://doi.org/10.1002/joc.1688

Daly, C., Neilson, R. P., & Phillips, D. L. (1994). A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *Journal of Applied Meteorology*, 33(2), 140–158. https://doi.org/10.1175/1520-0450(1994)033<0140:ASTMFM>2.0.CO;2

Dark, S. J., & Bram, D. (2007). The modifiable areal unit problem (MAUP) in physical geography. *Progress in Physical Geography*, 31(5), 471–479. https://doi.org/10.1177/0309133307083294

Das, D., Metzger, K., Heffernan, R., Balter, S., Weiss, D., & Mostashari, F. (2005). Monitoring Over-The-Counter Medication Sales for Early Detection of Disease Outbreaks --- New York City. Morbidity and Mortality Weekly Report (MMWR), 54(Suppl), 41–46. https://www.cdc.gov/mmwr/preview/mmwrhtml/su5401a9.htm

de Magny, G. C., Murtugudde, R., Sapiano, M. R. P., Nizam, A., Brown, C. W., Busalacchi, A. J., Yunus, M., Nair, G. B., Gil, A. I., Lanata, C. F., Calkins, J., Manna, B., Rajendran, K., Bhattacharya, M. K., Huq, A., Sack, R. B., & Colwell, R. R. (2008). Environmental signatures associated with cholera epidemics. *Proceedings of the National Academy of Sciences*, 105(46), 17676–17681. https://doi.org/10.1073/pnas.0809654105

de Man, H., Mughini Gras, L., Schimmer, B., Friesema, I. H. M., de Roda Husman, A. M., & van Pelt, W. (2016). Gastrointestinal, influenza-like illness and dermatological complaints following exposure to floodwater: a cross-sectional survey in The Netherlands. *Epidemiology and Infection*, 144(7), 1445–1454. https://doi.org/10.1017/S0950268815002654

de Roos, A. J., Kondo, M. C., Robinson, L. F., Rai, A., Ryan, M., Haas, C. N., Lojo, J., & Fagliano, J. A. (2020). Heavy precipitation, drinking water source, and acute gastrointestinal illness in Philadelphia, 2015-2017. PLOS ONE, 15(2), e0229258. https://doi.org/10.1371/journal.pone.0229258

de Schrijver, E., Folly, C. L., Schneider, R., Royé, D., Franco, O. H., Gasparrini, A., & Vicedo-Cabrera, A. M. (2021). A comparative analysis of the temperature-mortality risks using different weather datasets across heterogeneous regions. *GeoHealth*, 5(5), 1–15. https://doi.org/10.1029/2020GH000363

DeFelice, N. B. (2014). Drinking water risks to health 40 years after passage of the Safe Drinking Water Act: A county-by-county analysis in North Carolina [PhD Dissertation, University of North Carolina at Chapel Hill]. http://dissertations.umi.com/unc:14929

DeFelice, N. B., Johnston, J. E., & MacDonald Gibson, J. (2015). Acute gastrointestinal illness risks in North Carolina community water systems: A methodological comparison. *Environmental Science & Technology*, 49(16), 10019–10027. https://doi.org/10.1021/acs.est.5b01898

DeFelice, N. B., Johnston, J. E., & MacDonald Gibson, J. (2016). Reducing emergency department visits for acute gastrointestinal illnesses in in North Carolina (USA) by extending community water service. *Environmental Health Perspectives*, 124(10), 1583–1591. https://doi.org/10.1289/EHP160

Desai, R., Lopman, B. A., Shimshoni, Y., Harris, J. P., Patel, M. M., & Parashar, U. D. (2012). Use of internet search data to monitor impact of rotavirus vaccination in the United States. *Clinical Infectious Diseases*, 54(9), e115–e118. https://doi.org/10.1093/cid/cis121

Deschênes, O., & Greenstone, M. (2007). The Economic Impacts of Climate Change: Evidence from Agricultural Output and Random Fluctuations in Weather. *American Economic Review*, 97(1), 354–385. https://doi.org/10.1257/aer.97.1.354

Deshpande, A., Chang, H. H., & Levy, K. (2020). Heavy Rainfall Events and Diarrheal Diseases: The Role of Urban–Rural Geography. *The American Journal of Tropical Medicine and Hygiene*, 103(3), 1043–1049. https://doi.org/10.4269/ajtmh.19-0768

Dewan, A. M., Corner, R., Hashizume, M., & Ongee, E. T. (2013). Typhoid fever and its association with environmental factors in the Dhaka metropolitan area of Bangladesh: A spatial and time-series approach. *PLoS Neglected Tropical Diseases*, 7(1), e1998. https://doi.org/10.1371/journal.pntd.0001998

Dewan, S. (2005). In county made rich by golf, some enclaves are left behind. In The New York Times. https://www.nytimes.com/2005/06/07/us/in-county-made-rich-by-golf-some-enclaves-are-left-behind.html

Dhimal, M., Bhandari, D., Karki, K. B., Shrestha, S. L., Khanal, M., Shrestha, R. R. P., Dahal, S., Bista, B., Ebi, K. L., Cissé, G., Sapkota, A., & Groneberg, D. A. (2022). Effects of Climatic Factors on Diarrheal Diseases among Children below 5 Years of Age at National and Subnational Levels in Nepal: An Ecological Study. *International Journal of Environmental Research and Public Health*, 19(10), 6138. https://doi.org/10.3390/ijerph19106138

Dieter, C. A., Linsey, K. S., Caldwell, R. R., Harris, M. A., Ivahnenko, T. I., Lovelace, J. K., Maupin, M. A., & Barber, N. L. (2018). Estimated use of water in the United States county-level data for 2015 (ver. 2.0, June 2018). U.S. Geological Survey data release. https://doi.org/10.5066/F7TB15V5

Dieter, C. A., Maupin, M. A., Caldwell, R.R., Harris, M. A., Ivahnenko, T. I., Lovelace, J. K., Barber, N. L., & Linsey, K. S. (2018). Estimated use of water in the United States in 2015: U.S. Geological Survey Circular 1441. In US Geol Surv Circ (Issue 765). https://doi.org/10.3133/cir1441

Ding, G., Zhang, Y., Gao, L., Ma, W., Li, X., Liu, J., Liu, Q., & Jiang, B. (2013). Quantitative Analysis of Burden of Infectious Diarrhea Associated with Floods in Northwest of Anhui Province, China: A Mixed Method Evaluation. *PLoS ONE*, 8(6), e65112. https://doi.org/10.1371/journal.pone.0065112

Dixon, K. E. (1997). A comparison of case-crossover and case-control designs in a study of risk factors for hemorrhagic fever with renal syndrome. Epidemiology, 8(3), 243–246.

Dominici, F. (2004). Time-series analysis of air pollution and mortality: A statistical review. Research Report (Health Effects Institute), 123, 33. http://www.ncbi.nlm.nih.gov/pubmed/15757000

Dominici, F., Samet, J. M., & Zeger, S. L. (2000). Combining evidence on air pollution and daily mortality from the 20 largest US cities: a hierarchical modelling strategy. *Journal of the Royal Statistical Society: Series A* (Statistics in Society), 163(3), 263–302. https://doi.org/10.1111/1467-985X.00170

Dominici, F., Sheppard, L., & Clyde, M. (2003). Health effects of air pollution: A statistical review. *International Statistical Review*, 71, 243–276. https://doi.org/10.1111/j.1751-5823.2003.tb00195.x

Dominici, F., Zeger, S. L., & Samet, J. M. (2000). A measurement error model for time-series studies of air pollution and mortality. *Biostatistics*, 1(2), 157–175.

Drayna, P., McLellan, S. L., Simpson, P., Li, S.-H., & Gorelick, M. H. (2010). Association between rainfall and pediatric emergency department visits for acute gastrointestinal illness. *Environmental Health Perspectives*, 118(10), 1439–1443. https://doi.org/10.1289/ehp.0901671

D'Souza, R. M., Becker, N. G., Hall, G., & Moodie, K. B. A. (2004). Does Ambient Temperature Affect Foodborne Disease? *Epidemiology*, 15(1), 86–92. https://doi.org/10.1097/01.ede.0000101021.03453.3e

D'Souza, R. M., Hall, G., & Becker, N. G. (2008). Climatic factors associated with hospitalizations for rotavirus diarrhoea in children under 5 years of age. *Epidemiology and Infection*, 136(1), 56–64. https://doi.org/10.1017/S0950268807008229

Durman, C. F., Gregory, J. M., Hassell, D. C., Jones, R. G., & Murphy, J. M. (2001). A comparison of extreme European daily precipitation simulated by a global and a regional model for present and future climates. *Quarterly Journal of the Royal Meteorological Society*, 127(573), 1005–1015. https://doi.org/10.1256/smsqj.57315

Easterling, D. R., Arnold, J. R., Knutson, T., Kunkel, K. E., LeGrande, A. N., Leung, L. R., Vose, R. S., Waliser, D. E., & Wehner, M. F. (2017). Precipitation Change in the United States. Climate Science Special Report: Fourth National Climate Assessment, Volume I. In D. J. Wuebbles, D. W. Fahey, K. A. Hibbard, D. J. Dokken, B. C. Stewart, & T. K. Maycock (Eds.), Climate Science Special Report: Fourth National Climate Assessment, Volume I (pp. 207–230). https://doi.org/10.7930/J0H993CC

Eaves, L. A., Keil, A. P., Rager, J. E., George, A., & Fry, R. C. (2022). Analysis of the novel NCWELL database highlights two decades of co-occurrence of toxic metals in North Carolina private well water: Public health and environmental justice implications. *Science of the Total Environment*, 812, 151479. https://doi.org/10.1016/j.scitotenv.2021.151479

Egorov, A. I., Naumova, E. N., Tereschenko, A. A., Kislitsin, V. A., & Ford, T. E. (2003). Daily variations in effluent water turbidity and diarrhoeal illness in a Russian city. *International*

Journal of Environmental Health Research, 13(1), 81–94. https://doi.org/10.1080/0960312021000071567

Eisenberg, J. N. S., Desai, M. A., Levy, K., Bates, S. J., Liang, S., Naumoff, K., & Scott, J. C. (2007). Environmental determinants of infectious disease: A framework for tracking causal links and guiding public health research. *Environmental Health Perspectives*, 115(8), 1216–1223. https://doi.org/10.1289/ehp.9806

Eisenberg, J. N. S., Scott, J. C., & Porco, T. (2007). Integrating disease control strategies: balancing water sanitation and hygiene interventions to reduce diarrheal disease burden. *American Journal of Public Health*, 97(5), 846–852. https://doi.org/10.2105/AJPH.2006.086207

Eisenberg, J. N. S., Seto, E. Y. W., Colford, J. M., Olivieri, A., & Spear, R. C. (1998). An analysis of the Milwaukee cryptosporidiosis outbreak based on a dynamic model of the infection process. *Epidemiology*, 9(3), 255–263. http://www.jstor.org/stable/3703054

Eisenberg, J. N. S., Trostle, J., Sorensen, R. J. D., & Shields, K. F. (2012). Toward a Systems Approach to Enteric Pathogen Transmission: From Individual Independence to Community Interdependence. Annual Review of Public Health, 33(1), 239–257. https://doi.org/10.1146/annurev-publhealth-031811-124530

Eisenberg, M. C., Kujbida, G., Tuite, A. R., Fisman, D. N., & Tien, J. H. (2013). Examining rainfall and cholera dynamics in Haiti using statistical and dynamic modeling approaches. *Epidemics*, 5(4), 197–207. https://doi.org/10.1016/j.epidem.2013.09.004

Eisenhauer, I. F., Hoover, C. M., Remais, J. v., Monaghan, A., Celada, M., & Carlton, E. J. (2016). Estimating the risk of domestic water source contamination following precipitation events. *The American Journal of Tropical Medicine and Hygiene*, 94(6), 1403–1406. https://doi.org/10.4269/ajtmh.15-0600

El-Fadel, M., Ghanimeh, S., Maroun, R., & Alameddine, I. (2012). Climate change and temperature rise: Implications on food- and water-borne diseases. *Science of the Total Environment*, 437, 15–21. https://doi.org/10.1016/j.scitotenv.2012.07.041

Environmental Systems Research Institute (ESRI). (2013). Data and Maps for ArcGIS(R), 2013 [zip_usa.gdb; zip_poly.gdb]. ESRI, c2013.

Environmental Working Group (EWG), & Waterkeeper Alliance. (2016). Exposing Fields of Filth: Locations of Concentrated Animal Feeding Operations in North Carolina [dataset]. https://www.ewg.org/research/exposing-fields-filth

Epstein, A., Benmarhnia, T., & Weiser, S. D. (2020). Drought and illness among young Children in Uganda, 2009–2012. *The American Journal of Tropical Medicine and Hygiene*, 102(3), 644–648. https://doi.org/10.4269/ajtmh.19-0412

European Centre for Medium-Range Weather Forecasts (ECMWF). (2010). IFS Documentation - Cy36r1. Part II: Data assimilation. ECMWF.

Fang, X., Ai, J., Liu, W., Ji, H., Zhang, X., Peng, Z., Wu, Y., Shi, Y., Shen, W., & Bao, C. (2019). Epidemiology of infectious diarrhoea and the relationship with etiological and meteorological factors in Jiangsu Province, China. *Scientific Reports*, 9(1), 19571. https://doi.org/10.1038/s41598-019-56207-2

Febriani, Y., Levallois, P., Gingras, S., Gosselin, P., Majowicz, S. E., & Fleury, M. D. (2010). The association between farming activities, precipitation, and the risk of acute gastrointestinal illness in rural municipalities of Quebec, Canada: A cross-sectional study. *BMC Public Health*, 10(1), 48. https://doi.org/10.1186/1471-2458-10-48

Ferreira Braga, A. L., Zanobetti, A., & Schwartz, J. (2001). The time course of weather-related deaths. *Epidemiology*, 12(6), 662–667. https://doi.org/10.1097/00001648-200111000-00014

Fewtrell, L., Kaufmann, R. B., Kay, D., Enanoria, W., Haller, L., & Colford, J. M. (2005). Water, sanitation, and hygiene interventions to reduce diarrhoea in less developed countries: a systematic review and meta-analysis. *The Lancet Infectious Diseases*, 5(1), 42–52. https://doi.org/10.1016/S1473-3099(04)01253-8

Fewtrell, L., Kaufmann, R. B., Kay, D., Enanoria, W., Haller, L., & Colford Jr., J. M. (2005). Water, sanitation, and hygiene interventions to reduce diarrhoea in less developed countries: A systematic review and meta-analysis. *Lancet Infectious Diseases*, 5(1), 42–52. https://doi.org/10.1016/S1473-3099(04)01253-8

Fezzi, C., & Bateman, I. (2015). The impact of climate change on agriculture: Nonlinear effects and aggregation bias in ricardian models of farmland values. *Journal of the Association of Environmental and Resource Economist*s, 2(1), 57–92. https://doi.org/10.1086/680257

Finch, W. H., Bolin, J. E., & Kelley, K. (2019). Multilevel Modeling Using R (2nd ed.). Chapman & Hall/CRC.

Finger, F., Knox, A., Bertuzzo, E., Mari, L., Bompangue, D., Gatto, M., Rodriguez-Iturbe, I., & Rinaldo, A. (2014). Cholera in the Lake Kivu region (DRC): Integrating remote sensing and spatially explicit epidemiological modeling. *Water Resources Research*, 50(7), 5624–5637. https://doi.org/10.1002/2014WR015521

Fink, G., Gunther, I., & Hill, K. (2011). The effect of water and sanitation on child health: evidence from the demographic and health surveys 1986-2007. *International Journal of Epidemiology*, 40(5), 1196–1204.

Fischer Walker, C. L., Aryee, M. J., Boschi-Pinto, C., & Black, R. E. (2012). Estimating diarrhea mortality among young children in low and middle income countries. *PLoS ONE*, 7(1), 1–7. https://doi.org/10.1371/journal.pone.0029151

Fischer Walker, C. L., Perin, J., Aryee, M. J., Boschi-Pinto, C., & Black, R. E. (2012). Diarrhea incidence in low- and middle-income countries in 1990 and 2010: A systematic review. *BMC Public Health*, 12(1). https://doi.org/10.1186/1471-2458-12-220

Fisher, A. C., Hanemann, W. M., Roberts, M. J., & Schlenker, W. (2012). The Economic Impacts of Climate Change: Evidence from Agricultural Output and Random Fluctuations in Weather: Comment. *American Economic Review*, 102(7), 3749–3760.

Fisman, D. N. (2007). Seasonality of infectious diseases. *Annual Review of Public Health*, 28(1), 127–143. https://doi.org/10.1146/annurev.publhealth.28.021406.144128

Fisman, D. N., Lim, S., Wellenius, G. A., Johnson, C., Britz, P., Gaskins, M., Maher, J., Mittleman, M. A., Spain, C. V., Haas, C. N., Newbern, C., Spain, C. V., Haas, C. N., & Newbern, C. (2005). It's not the heat, it's the humidity: Wet weather increases legionellosis risk in the greater Philadelphia metropolitan area. *The Journal of Infectious Diseases*, 192(12), 2066–2073. https://doi.org/10.1086/498248

Flahault, A., & Hanslik, T. (2010). Epidemiology of viral gastroenteritis in France and Europe. Bulletin de l'Academie Nationale de Medecine, 194(8), 1415–1425. https://doi.org/10.1016/s0001-4079(19)32172-7

Flanagan, S. v., Spayd, S. E., Procopio, N. A., Marvinney, R. G., Smith, A. E., Chillrud, S. N., Braman, S., & Zheng, Y. (2016). Arsenic in private well water part 3 of 3: Socioeconomic vulnerability to exposure in Maine and New Jersey. *Science of the Total Environment*, 562, 1019–1030. https://doi.org/10.1016/j.scitotenv.2016.03.217

Fleury, M., Charron, D. F., Holt, J. D., Allen, O. B., & Maarouf, A. R. (2006). A time series analysis of the relationship of ambient temperature and common bacterial enteric infections in two Canadian provinces. *International Journal of Biometeorology*, 50(6), 385–391. https://doi.org/10.1007/s00484-006-0028-9

Fotheringham, A. S., & Wong, D. W. S. (1991). The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A*, 23(7), 1025–1044. https://doi.org/10.1068/a231025

Frei, C., & Schär, C. (2001). Detection probability of trends in rare events: Theory and application to heavy precipitation in the Alpine region. *Journal of Climate*, 14(7), 1568–1584. https://doi.org/10.1175/1520-0442(2001)014<1568:DPOTIR>2.0.CO;2

Friedman, J. H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics*, 19(1), 1–67. https://doi.org/10.1214/aos/1176347963

Fuente, D., Allaire, M., Jeuland, M., & Whittington, D. (2020). Forecasts of mortality and economic losses from poor water and sanitation in sub-Saharan Africa. *PLOS ONE*, 15(3), e0227611. https://doi.org/10.1371/journal.pone.0227611

Galway, L. P., Allen, D. M., Parkes, M. W., Li, L., & Takaro, T. K. (2015). Hydroclimatic variables and acute gastro-intestinal illness in British Columbia, Canada: A time series analysis. *Water Resources Research*, 51(2), 885–895. https://doi.org/10.1002/2014WR015519

Gao, L., Zhang, Y., Ding, G., Liu, Q., Zhou, M., Li, X., & Jiang, B. (2014). Meteorological Variables and Bacillary Dysentery Cases in Changsha City, China. *The American Journal of Tropical Medicine and Hygien*e, 90(4), 697–704. https://doi.org/10.4269/ajtmh.13-0198

Gao, Y., Chen, Y., Shi, P., Zhang, Q., Qian, C., Xiao, Y., Feng, W., Shen, Y., & Shi, C. (2020). The effect of ambient temperature on infectious diarrhea and diarrhea-like illness in Wuxi, China. Disaster Medicine and Public Health Preparedness, 16(2), 583–589. https://doi.org/10.1017/dmp.2020.340

Gardner, W. A. (1985). Introduction to Random Processes with Applications to Signals and Systems. Macmillan.

Gardner, W. A., Napolitano, A., & Paura, L. (2006). Cyclostationarity: Half a century of research. Signal Processing, 86(4), 639–697. https://doi.org/10.1016/j.sigpro.2005.06.016

Garthright, W. E., Archer, D. L., & Kvenberg, J. E. (1988). Estimates of incidence and costs of intestinal infectious diseases in the United States. *Public Health Reports*, 103(2), 107–115. /pmc/articles/PMC1477958/?report=abstract

Gasparrini, A. (2011). Distributed lag linear and non-linear models in R: The package dlnm. Journal of Statistical Software, 43(8), 1–20. http://www.lshtm.ac.uk/people/gasparrini.antonio/.

Gasparrini, A. (2014). Modeling exposure-lag-response associations with distributed lag non-linear models. *Statistics in Medicine*, 33(5), 881–899. https://doi.org/10.1002/sim.5963

Gasparrini, A., & Armstrong, B. (2010). Time series analysis on the health effects of temperature: Advancements and limitations. *Environmental Research*, 110(6), 633–638. https://doi.org/10.1016/j.envres.2010.06.005

Gasparrini, A., Armstrong, B., & Kenward, M. G. (2010). Distributed lag non-linear models. *Statistics in Medicine*, 29(21), 2224–2234. https://doi.org/10.1002/sim.3940

Gasparrini, A., & Leone, M. (2014). Attributable risk from distributed lag models. *BMC Medical Research Methodology*, 14(1), 55. https://doi.org/10.1186/1471-2288-14-55

Gasparrini, A., Scheipl, F., Armstrong, B., & Kenward, M. G. (2017). A penalized framework for distributed lag non-linear models. *Biometrics*, 73(3), 938–948. https://doi.org/10.1111/biom.12645

Gelman, A., & Hill, J. (2008). Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press.

Gelman, A., Park, D. K., Ansolabehere, S., Price, P. N., & Minnite, L. C. (2001). Models, assumptions and model checking in ecological regressions. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, 164(1), 101–118. https://doi.org/10.1111/1467-985X.00190

Gibbons, C. L., Mangen, M.-J. J., Plass, D., Havelaar, A. H., Brooke, R. J., Kramarz, P., Peterson, K. L., Stuurman, A. L., Cassini, A., Fèvre, E. M., & Kretzschmar, M. E. (2014). Measuring underreporting and under-ascertainment in infectious disease datasets: a comparison of methods. *BMC Public Health*, 14(1), 147. https://doi.org/10.1186/1471-2458-14-147

Gilbert, P. (2013). The state of exclusion: An empirical analysis of the legacy of segregated communities in North Carolina. UNC Center for Civil Rights. http://www.uncinclusionproject.org/documents/stateofexclusion.pdf

Gilliam, J. W., Osmond, D. L., & Evans, R. O. (1997). Selected agricultural best management practices to control nitrogen in the Neuse River Basin. In North Carolina Agricultural Research Service Technical Bulletin (Vol. 311). N.C. Agricultural Research Service, N.C. State University. https://agris.fao.org/agris-search/search.do?recordID=US201300042644

Giorgi, F., Coppola, E., & Raffaele, F. (2014). A consistent picture of the hydroclimatic response to global warming from multiple indices: Models and observations. *Journal of Geophysical Research: Atmospheres*, 119(20), 11,695-11,708. https://doi.org/10.1002/2014JD022238

Gleason, J. A., & Fagliano, J. A. (2017). Effect of drinking water source on associations between gastrointestinal illness and heavy rainfall in New Jersey. *PLoS ONE*, 12(3), 1-13 (e0173794). https://doi.org/10.1371/journal.pone.0173794

Goldberg, M. S., Burnett, R. T., & Stieb, D. (2003). A review of time-series studies used to evaluate the short-term effects of air pollution on human health. *Reviews on Environmental Health*, 18(4), 269–303. https://doi.org/10.1515/REVEH.2003.18.4.269

Gosai, A., Salinger, J., & Dirks, K. (2009). Climate and respiratory disease in Auckland, New Zealand. *Australian and New Zealand Journal of Public Health*, 33(6), 521–526. https://doi.org/10.1111/j.1753-6405.2009.00447.x

Graddy, S., Simon, E., & Rundquist, S. (2020). UPDATE: Exposing Fields of Filth: Factory Farms Disproportionately Threaten Black, Latino and Native American North Carolinians.

Grasso, M., Manera, M., Chiabai, A., & Markandya, A. (2012). The health effects of climate change: A survey of recent quantitative research. *International Journal of Environmental Research and Public Health*, 9(5), 1523–1547. https://doi.org/10.3390/ijerph9051523

Graydon, R. C., Mezzacapo, M., Boehme, J., Foldy, S., Edge, T. A., Brubacher, J., Chan, H. M., Dellinger, M., Faustman, E. M., Rose, J. B., & Takaro, T. K. (2022). Associations between extreme precipitation, drinking water, and protozoan acute gastrointestinal illnesses in four North American Great Lakes cities (2009–2014). *Journal of Water and Health*, 20(5), 849–862. https://doi.org/10.2166/wh.2022.018

Greer, A., Ng, V., & Fisman, D. (2008). Climate change and infectious diseases in North America: the road ahead. *Canadian Medical Association Journal*, 178(6), 715–722.

Grjibovski, A. M., Bushueva, V., Boltenkov, V. P., Buzinov, R. V, Degteva, G. N., Yurasova, E. D., & Nurse, J. (2013). Climate variations and salmonellosis in northwest Russia: a time-series analysis. *Epidemiology & Infection,* 141(02), 269–276.

Grjibovski, A. M., Kosbayeva, A., & Menne, B. (2014). The effect of ambient air temperature and precipitation on monthly counts of salmonellosis in four regions of Kazakhstan, Central Asia, in 2000–2010. *Epidemiology and Infection*, 142(3), 608–615. https://doi.org/10.1017/S095026881300157X

Grubesic, T. H., & Matisziw, T. C. (2006). On the use of ZIP codes and ZIP code tabulation areas (ZCTAs) for the spatial analysis of epidemiological data. *International Journal of Health Geographics*, 5(1), 58. https://doi.org/10.1186/1476-072X-5-58

Guan, P., Huang, D., Guo, J., Wang, P., & Zhou, B. (2008). Bacillary dysentery and meteorological factors in northeastern China: a historical review based on classification and regression trees. *Japanese Journal of Infectious Diseases*, 61(5), 356–360.

Gunnarsdottir, M. J., Gardarsson, S. M., & Andradottir, H. O. (2013). Microbial contamination in groundwater supply in a cold climate and coarse soil: Case study of norovirus outbreak at Lake Mývatn, Iceland. *Hydrology Research*, 44(6), 1114–1128. https://doi.org/10.2166/nh.2013.076

Guzman Herrador, B. R., de Blasio, B. F., MacDonald, E., Nichols, G., Sudre, B., Vold, L., Semenza, J. C., & Nygård, K. (2015). Analytical studies assessing the association between extreme precipitation or temperature and drinking water-related waterborne infections: a review. *Environmental Health*, 14(29), 1–12. https://doi.org/10.1186/s12940-015-0014-y

Haffejee, I. E. (1995). The epidemiology of rotavirus infections: A global perspective. *Journal of Pediatric Gastroenterology and Nutrition*, 20(3), 275–286. https://doi.org/10.1097/00005176-199504000-00003

Haines, A., & Patz, J. A. (2004). Health effects of climate change. *Journal of the American Medical Association*, 291(1), 3–99.

Hakenwerth, A. M., Waller, A. E., Tintinalli, J., & Ising, A. (2009). North Carolina Disease Event Tracking and Epidemiologic Collection Tool (NC-DETECT) Annual Report. North Carolina Emergency Department Data January 1, 2007-December 31, 2007. http://ncdetect.org/pubs.html

Halekoh, U., Højsgaard, S., & Yan, J. (2006). The R package geepack for generalized estimating equations. *Journal of Statistical Software*, 15(2), 1–11. https://doi.org/10.18637/jss.v015.i02

Hales, S., Kovats, S., Lloyd, S., & Campbell-Lendrum, D. (2014). Quantitative risk assessment of the effects of climate change on selected causes of death, 2030s and 2050s (S. Hales, S. Kovats, S. Lloyd, & D. Campbell-Lendrum, Eds.). World Health Organization (WHO). http://www.who.int/globalchange/publications/quantitative-risk-assessment/en/

Hall, G. V., Hanigan, I. C., Dear, K. B. G., & Vally, H. (2011). The influence of weather on community gastroenteritis in Australia. *Epidemiology and Infection*, 139(6), 927–936. https://doi.org/10.1017/S0950268810001901

Hanemann, W. M., Jeuland, M. A., Downs, K. D., & Whittington, D. (2014). Estimating the global health damages from climate change in an Integrated Assessment Model: The problems of scale and aggregation. In World Congress for Environmental and Resource Economists, 29 June – 2 July 2014.

Hao, Y., Liao, W., Ma, W., Zhang, J., Zhang, N., Zhong, S., Wang, Z., Yang, L., & Huang, C. (2019). Effects of ambient temperature on bacillary dysentery: A multi-city analysis in Anhui Province, China. *Science of the Total Environment*, 671, 1206–1213. https://doi.org/10.1016/j.scitotenv.2019.03.443

Hardy, S. (2012). The price of pork [WWW Document]. UNC Endeavors. http://endeavors.unc.edu/the_price_of_pork

Hartley, J. M. (2016). Spatial and temporal patterns of gastrointestinal illness and their relationship with precipitation across the state of North Carolina [Master's Thesis] [Master's Thesis]. University of North Carolina at Chapel Hill.

Hashizume, M., Armstrong, B. G., Hajat, S., Wagatsuma, Y., Faruque, A. S. G., Hayashi, T., & Sack, D. A. (2007a). Association between climate variability and hospital visits for non-cholera diarrhoea in Bangladesh: effects and vulnerable groups. *International Journal of Epidemiology*, 36(5), 1030–1037. https://doi.org/10.1093/ije/dym148

Hashizume, M., Armstrong, B. G., Wagatsuma, Y., Faruque, A. S. G., Hayashi, T., Sack, D. A. A., & Faruque, G. (2008a). Rotavirus infections and climate variability in Dhaka, Bangladesh: A time-series analysis. *Epidemiology & Infection*, 136(9), 1281–1289. https://doi.org/10.1017/S0950268807009776

Hashizume, M., Armstrong, B. G., Wagatsuma, Y., Faruque, A. S. G., Hayashi, T., Sack, D. A. A., & Faruque, G. (2008b). Rotavirus infections and climate variability in Dhaka, Bangladesh: A time-series analysis. *Epidemiology & Infection*, 136(9), 1281–1289. https://doi.org/10.1017/S0950268807009776

Hashizume, M., Armstrong, B., Hajat, S., Wagatsuma, Y., Faruque, A. S. G., Hayashi, T., & Sack, D. A. (2008a). The effect of rainfall on the incidence of cholera in Bangladesh. *Epidemiology*, 19(1), 103–110. https://doi.org/10.1097/EDE.0b013e31815c09ea

Hashizume, M., Armstrong, B., Hajat, S., Wagatsuma, Y., Faruque, A. S. G., Hayashi, T., & Sack, D. A. (2008b). The Effect of Rainfall on the Incidence of Cholera in Bangladesh. Epidemiology, 19(1), 103–110. https://doi.org/10.1097/EDE.0b013e31815c09ea

Hashizume, M., Armstrong, B., Hajat, S., Wagatsuma, Y., Faruque, A. S., Hayashi, T., & Sack, D. A. (2007b). Association between climate variability and hospital visits for non-cholera diarrhoea in Bangladesh: effects and vulnerable groups. *International Journal of Epidemiology*, 36(5), 1030–1037. https://doi.org/10.1093/ije/dym148

Hashizume, M., Chaves, L. F., Faruque, A. S. G., Yunus, M., Streatfield, K., & Moji, K. (2013). A differential effect of Indian Ocean dipole and El Niño on cholera dynamics in Bangladesh. *PLoS ONE,* 8(3), e60001. https://doi.org/10.1371/journal.pone.0060001

Hashizume, M., Faruque, A. S. G., Terao, T., Yunus, M., Streatfield, K., Yamamoto, T., & Moji, K. (2011a). The Indian Ocean dipole and cholera incidence in Bangladesh: A time-series analysis. *Environmental Health Perspectives*, 119(2), 239–244. https://doi.org/10.1289/ehp.1002302

Hashizume, M., Faruque, A. S. G., Terao, T., Yunus, M., Streatfield, K., Yamamoto, T., & Moji, K. (2011b). The indian ocean dipole and cholera incidence in Bangladesh: A time-series analysis. *Environmental Health Perspectives*, 119(2), 239–244. https://doi.org/10.1289/ehp.1002302

Hashizume, M., Wagatsuma, Y., Faruque, A. S. G., Sack, D. A., Hayashi, T., Hunter, P. R., Armstrong, B. G., & Sack, D. A. (2008). Factors determining vulnerability to diarrhoea during and after severe floods in Bangladesh. *Journal of Water and Health*, 6(3), 323–332. https://doi.org/10.2166/wh.2008.062

Hastie, T. (2020). Package "gam." CRAN.

Hastie, T., & Tibshirani, R. (1986). Generalized additive models. *Statistical Science*, 1(3), 297–310. https://doi.org/10.1214/ss/1177013604

Hastie, T., & Tibshirani, R. (1990a). *Generalized Additive Models (1st ed.).* Chapman and Hall.

Hastie, T., & Tibshirani, R. (1990b). Exploring the nature of covariate effects in the proportional hazards model. *Biometric*s, 46(4), 1005. https://doi.org/10.2307/2532444

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction (1st ed.)*. Springer. https://doi.org/10.1007/978-0-387-84858-7

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.)*. https://doi.org/10.1007/978-0-387-84858-7

He, S., Mazumdar, S., & Arena, V. C. (2006). A comparative study of the use of GAM and GLM in air pollution research. *Environmetrics*, 17(1), 81–93. https://doi.org/10.1002/env.751

Herrera, D., Ellis, A., Fisher, B., Golden, C. D., Johnson, K., Mulligan, M., Pfaff, A., Treuer, T., & Ricketts, T. H. (2017). Upstream watershed condition predicts rural children's health across 35 developing countries. *Nature Communications*, 8(1), 811. https://doi.org/10.1038/s41467-017-00775-2

Herrick, K. A., Huettmann, F., & Lindgren, M. A. (2013). A global model of avian influenza prediction in wild birds: the importance of northern regions. *Veterinary Research*, 44(1), 42. https://doi.org/10.1186/1297-9716-44-42

Hervás, D., Hervás-Masip, J., Rosell, A., Mena, A., Pérez, J. L., & Hervás, J. A. (2014). Are hospitalizations for rotavirus gastroenteritis associated with meteorologic factors? *European Journal of Clinical Microbiology & Infectious Diseases*, 33(9), 1547–1553. https://doi.org/10.1007/s10096-014-2106-y

Heywood, D. I., Cornelius, S., & Carver, S. (1988). An Introduction to Geographical Information Systems. Addison Wesley Longman.

Hock, R., Rasul, G., Adler, C., Cáceres, B., Gruber, S., Hirabayashi, Y., Jackson, M., Kääb, A., Kang, S., Kutuzov, S., Milner, A., Molau, U., Morin, S., Orlove, B., & Steltzer, H. I. (2022). High Mountain Areas. In H.-O. Pörtner, D. C. Roberts, V. Masson-Delmotte, P. Zhai, M. Tignor, E. Poloczanska, K. Mintenbeck, A. Alegría, M. Nicolai, A. Okem, J. Petzold, B. Rama, & N. M. Weyer (Eds.), IPCC Special Report on the Ocean and Cryosphere in a Changing Climate (pp. 131–202). Cambridge University Press. https://doi.org/10.1017/9781009157964.004

Hofstra, N., Haylock, M., New, M., Jones, P., & Frei, C. (2008). Comparison of six methods for the interpolation of daily, European climate data. *Journal of Geophysical Research Atmospheres*, 113(21). https://doi.org/10.1029/2008JD010100

Holcomb, D. A., Quist, A. J. L., & Engel, L. S. (2022). Exposure to industrial hog and poultry operations and urinary tract infections in North Carolina, USA. *Science of the Total Environment*, 853, 158749. https://doi.org/10.1016/j.scitotenv.2022.158749

Hope, C. W. (2011). The Social Cost of CO2 from the PAGE09 Model. Economics Discussion Papers, No 2011-39, Kiel Institute for the World Economy, 2011–39, Art. 2011–39. https://doi.org/10.2139/ssrn.1973863

Hope, C. W. (2013). Critical issues for the calculation of the social cost of CO2: why the estimates from PAGE09 are higher than those from PAGE2002. *Climatic Change*, 117(3), 531–543. https://doi.org/10.1007/s10584-012-0633-z

Horn, L., Hajat, A., Sheppard, L., Quinn, C., Colborn, J., Zermoglio, M., Gudo, E., Marrufo, T., & Ebi, K. (2018). Association between precipitation and diarrheal disease in Mozambique. *International Journal of Environmental Research and Public Health*, 15(4), 709. https://doi.org/10.3390/ijerph15040709

Hribar, C. (2010). Understanding concentrated animal feeding operations and their impact on communities [Internet]. 1–22. http://www.cdc.gov/nceh/ehs/docs/understanding_cafos_nalboh.pdf

Hsieh, J. L., Nguyen, T. Q., Matte, T., & Ito, K. (2015). Drinking water turbidity and emergency department visits for gastrointestinal illness in New York City, 2002-2009. *PLOS ONE,* 10(4), e0125071. https://doi.org/10.1371/journal.pone.0125071

Hu, W., Mengersen, K., Fu, S.-Y., & Tong, S. (2010). The use of ZIP and CART to model cryptosporidiosis in relation to climatic variables. *International Journal of Biometeorology*, 54(4), 433–440. https://doi.org/10.1007/s00484-009-0294-4

Hu, W., Mengersen, K., & Tong, S. (2010). Risk factor analysis and spatiotemporal CART model of cryptosporidiosis in Queensland, Australia. *BMC Infectious Diseases*, 10(1), 311. https://doi.org/10.1186/1471-2334-10-311

Hu, W., Tong, S., Mengersen, K., & Connell, D. (2007). Weather variability and the incidence of cryptosporidiosis: Comparison of time series poisson regression and SARIMA models. *Annals of Epidemiology*, 17(9), 679–688. https://doi.org/10.1016/j.annepidem.2007.03.020

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., & Tan, J. (2019). GPM IMERG Final Precipitation L3 Half Hourly 0.1 degree x 0.1 degree V06. Goddard Earth Sciences Data and Information Services Center (GES DISC). https://doi.org/10.5067/GPM/IMERG/3B-HH/0

Hunter, P. R., & Wang, X. (2010). A Systematic Review and Meta-Analysis of the Association between Self-Reported Diarrheal Disease and Distance from Home to Water Source. *The American Journal of Tropical Medicine and Hygiene*, 83(3), 582–584. https://doi.org/10.4269/ajtmh.2010.10-0215

Huq, A., Sack, R. B., Nizam, A., Longini, I. M., Nair, G. B., Ali, A., Morris, J. G., Khan, M. N. H., Siddique, A. K., Yunus, M., Albert, M. J., Sack, D. A., & Colwell, R. R. (2005). Critical factors influencing the occurrence of Vibrio cholerae in the environment of Bangladesh. *Applied and Environmental Microbiology,* 71(8), 4645–4654. https://doi.org/10.1128/AEM.71.8.4645-4654.2005

Hutton, G., & Haller, L. (1994). Evaluation of the Costs and Benefits of Water and Sanitation Improvements at the Global Level. https://apps.who.int/iris/bitstream/handle/10665/68568/WHO_SDE_WSH_04.04.pdf

ICIMOD. (2010). Climate change vulnerability of mountain ecosystems in the Eastern Himalayas (Karma Tse-ring, E. Sharma, N. Chettri, & A. Shrestha, Eds.). International Centre for Integrated Mountain Development. https://doi.org/10.4000/anthrovision.589

Imai, C., Armstrong, B., Chalabi, Z., Mangtani, P., & Hashizume, M. (2015). Time series regression model for infectious disease and weather. *Environmental Research*, 142, 319–327. https://doi.org/10.1016/j.envres.2015.06.040

Imai, C., & Hashizume, M. (2015). A systematic review of methodology: Time series regression analysis for environmental factors and infectious diseases. *Tropical Medicine and Health*, 43(1), 1–9. https://doi.org/10.2149/tmh.2014-21

IPCC. (2013). Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, & P. M. Midgley, Eds.). Cambridge University Press.

Jagai, J. S., DeFlorio-Barker, S., Lin, C. J., Hilborn, E. D., & Wade, T. J. (2017). Sanitary sewer overflows and emergency room visits for gastrointestinal illness: Analysis of Massachusetts data, 2006-2007. *Environmental Health Perspectives*, 125(11), 117007. https://doi.org/10.1289/EHP2048

Jagai, J. S., Griffiths, J. K., Kirshen, P. K., Webb, P., & Naumova, E. N. (2012). Seasonal patterns of gastrointestinal illness and streamflow along the Ohio River. *International Journal of Environmental Research and Public Health*, 9(5), 1771–1790. https://doi.org/10.3390/ijerph9051771

Jagai, J. S., Li, Q., Wang, S., Messier, K. P., Wade, T. J., & Hilborn, E. D. (2015). Extreme precipitation and emergency room visits for gastrointestinal illness in areas with and without combined sewer systems: An analysis of Massachusetts data, 2003–2007. *Environmental Health Perspectives*, 123(9), 873–879. https://doi.org/10.1289/ehp.1408971

Jagai, J. S., Sarkar, R., Castronovo, D., Kattula, D., McEntee, J., Ward, H., Kang, G., & Naumova, E. N. (2012). Seasonality of rotavirus in South Asia: A meta-analysis approach assessing associations with temperature, precipitation, and vegetation index. *PLoS ONE*, 7(5), e38168. https://doi.org/10.1371/journal.pone.0038168

Jeuland, M. A., Fuente, D. E., Ozdemir, S., Allaire, M. C., & Whittington, D. (2013). The Long-Term Dynamics of Mortality Benefits from Improved Water and Sanitation in Less Developed Countries. *PLoS ONE,* 8(10), e74804. http://dx.doi.org/10.1371/journal.pone.0074804

Jiang, C., Shaw, K. S., Upperman, C. R., Blythe, D., Mitchell, C., Murtugudde, R., Sapkota, A. R., & Sapkota, A. (2015). Climate change, extreme events and increased risk of salmonellosis in Maryland, USA: Evidence for coastal vulnerability. *Environment International*, 83, 58–62. https://doi.org/10.1016/j.envint.2015.06.006

Johnson, J. H., Parnell, A., Joyner, A. M., Christman, C. J., & Marsh, B. (2004). Racial apartheid in a small North Carolina town. *The Review of Black Political Economy*, 31(4), 89–107. https://doi.org/10.1007/s12114-004-1012-4

Johnson, T. D., & Belitz, K. (2017). Domestic well locations and populations served in the conterminous U.S.:1990. U.S. Geological Survey data release. https://doi.org/10.5066/F7028PSX

Johnson, T. D., Belitz, K., & Lombard, M. A. (2019). Estimating domestic well locations and populations served in the contiguous U.S. for years 2000 and 2010. *Science of the Total Environment,* 687, 1261–1273. https://doi.org/10.1016/j.scitotenv.2019.06.036

Jones, T. F., McMillian, M. B., Scallan, E., Frenzen, P. D., Cronquist, A. B., Thomas, S., & Angulo, F. J. (2007). A population-based estimate of the substantial burden of diarrhoeal disease in the United States; FoodNet, 1996–2003. *Epidemiology and Infection*, 135(2), 293–301. https://doi.org/10.1017/S0950268806006765

Joyce, R. J., Janowiak, J. E., Arkin, P. A., & Xie, P. (2004). CMORPH: A Method that Produces Global Precipitation Estimates from Passive Microwave and Infrared Data at High Spatial and Temporal Resolution. *Journal of Hydrometeorology*, 5(3), 487–503. https://doi.org/10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2

Julian, T. R. (2016). Environmental transmission of diarrheal pathogens in low and middle income countries. In Environmental Science: Processes and Impacts (Vol. 18, Issue 8, pp. 944–955). Royal Society of Chemistry. https://doi.org/10.1039/c6em00222f

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, B., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., … Joseph, D. (1996). The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of the American Meteorological Society,* March, 437–472. https://doi.org/https://journals.ametsoc.org/view/journals/bams/77/3/1520-0477_1996_077_0437_tnyrp_2_0_co_2.xml

Kane, M. J., Price, N., Scotch, M., & Rabinowitz, P. (2014). Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks. *BMC Bioinformatics*, 15(1), 1–9. https://doi.org/10.1186/1471-2105-15-276

Kang, R., Xun, H., Zhang, Y., Wang, W., Wang, X., Jiang, B., & Ma, W. (2015). Impacts of different grades of tropical cyclones on infectious diarrhea in Guangdong, 2005-2011. *PLOS ONE*, 10(6), e0131423. https://doi.org/10.1371/journal.pone.0131423

Kaper, J. B., Nataro, J. P., & Mobley, H. L. T. (2004). Pathogenic *Escherichia coli*. *Nature Reviews Microbiology*, 2(2), 123–140. https://doi.org/10.1038/nrmicro818

Karl, T. R., & Knight, R. W. (1998). Secular trends of precipitation amount, frequency, and intensity in the United States. *Bulletin of the American Meteorological Society*, 79(2), 231–241. https://doi.org/10.1175/1520-0477(1998)079<0231:STOPAF>2.0.CO;2

Karlsson, S. (2013). Forecasting with Bayesian Vector Autoregression. In *Handbook of Economic Forecasting* (Vol. 2, Issue Part B, pp. 791–897). https://doi.org/10.1016/B978-0-444-62731-5.00015-4

Kay, D., & Falconer, R. (2008). Hydro-epidemiology: the emergence of a research agenda. Environmental Fluid Mechanics, 8(5–6), 451–459.

Kelly-Hope, L., & Thomson, M. C. (2008). Climate and Infectious Diseases. In *Seasonal Forecasts, Climatic Change and Human Health* (Vol. 30, pp. 31–70). Springer Netherlands.

Khormi, H. M., & Kumar, L. (2012). The importance of appropriate temporal and spatial scales for dengue fever control and management. *Science of the Total Environment*, 430(0), 144–149. https://doi.org/http://dx.doi.org/10.1016/j.scitotenv.2012.05.001

Kistler, R., Kalnay, E., Collins, W., Saha, S., White, G., Woollen, J., Chelliah, M., Ebisuzaki, W., Kanamitsu, M., Kousky, V., van den Dool, H., Jenne, R., & Fiorino, M. (2001). The NCEP-

NCAR 50-year reanalysis: Monthly means CD-ROM and documentation. *Bulletin of the American Meteorological Society*, 82(2), 247–267. https://doi.org/10.1175/1520-0477(2001)082<0247:TNNYRM>2.3.CO;2

Klein Tank, A. M. G., & Können, G. P. (2003). Trends in indices of daily temperature and precipitation extremes in Europe, 1946–99. *Journal of Climate*, 16(22), 3665–3680. https://doi.org/10.1175/1520-0442(2003)016<3665:TIIODT>2.0.CO;2

Koelle, K. (2009). The impact of climate on the disease dynamics of cholera. *Clinical Microbiology and Infection*, 15, 29–31. https://doi.org/10.1111/j.1469-0691.2008.02686.x

Koelle, K., & Pascual, M. (2004). Disentangling extrinsic from intrinsic factors in disease dynamics: A nonlinear time series approach with an application to cholera. *American Naturalist*, 163(6), 901–913. https://doi.org/10.1086/420798

Koelle, K., Rodó, X., Pascual, M., Yunus, M., & Mostafa, G. (2005). Refractory periods and climate forcing in cholera dynamics. *Nature*, 436(7051), 696–700. https://doi.org/10.1038/nature03820

Kolstad, E. W., & Johansson, K. A. (2011). Uncertainties associated with quantifying climate change impacts on human health: A case study for diarrhea. *Environmental Health Perspectives*, 119(3), 299–305. https://doi.org/10.1289/ehp.1002060

Koopman, J. S., & Longini Jr, I. M. (1994). The ecological effects of individual exposures and nonlinear disease dynamics in populations. *American Journal of Public Health*, 84(5), 836–842.

Kotloff, K. L., Nataro, J. P., Blackwelder, W. C., & Nasrin, D. (2013). Burden and aetiology of diarrhoeal disease in infants and young children in developing countries. *The Lancet*, 382(9888), 209.

Kotloff, K. L., Nataro, J. P., Blackwelder, W. C., Nasrin, D., Farag, T. H., Panchalingam, S., Wu, Y., Sow, S. O., Sur, D., Breiman, R. F., Faruque, A. S., Zaidi, A. K., Saha, D., Alonso, P. L., Tamboura, B., Sanogo, D., Onwuchekwa, U., Manna, B., Ramamurthy, T., … Levine, M. M. (2013). Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. *The Lancet*, 382(9888), 209–222. https://doi.org/10.1016/S0140-6736(13)60844-2

Kottek, M., Grieser, J., Beck, C., Rudolf, B., & Rubel, F. (2006). Map of the Köppen-Geiger climate classification updated. Meteorologische Zeitschrift, 15(3), 259–263. https://doi.org/10.1127/0941-2948/2006/0130

Kovats, R. S., Bouma, M. J., Hajat, S., Worrall, E., & Haines, A. (2003). El Niño and health. *The Lancet,* 362(9394), 1481–1489. https://doi.org/10.1016/S0140-6736(03)14695-8

Kovats, R. S., Edwards, S. J., Hajat, S., Armstrong, B. G., Ebi, K. L., Menne, B., Cowden, J., Gerner-Smidt, P., Hernández Pezzi, G., Kristufkova, Z., Kriz, B., Kutsar, K., Magdzik, W., O'Brien,

S. J., Schmid, H., & van Pelth, W. (2004). The effect of temperature on food poisoning: A time-series analysis of salmonellosis in ten European countries. *Epidemiology and Infection*, 132(3), 443–453. https://doi.org/10.1017/S0950268804001992

Kowalzik, F., Riera-Montes, M., Verstraeten, T., & Zepp, F. (2015). The Burden of Norovirus Disease in Children in the European Union. *The Pediatric Infectious Disease Journal*, 34(3), 229–234. https://doi.org/10.1097/INF.0000000000000546

Kraay, A. N. M., Man, O., Levy, M. C., Levy, K., Ionides, E., & Eisenberg, J. N. S. (2020). Understanding the impact of rainfall on diarrhea: Testing the concentration-dilution hypothesis using a systematic review and meta-analysis. *Environmental Health Perspectives*, 128(12), 126001. https://doi.org/10.1289/EHP6181

Krieger, N., Waterman, P., Chen, J. T., Soobader, M.-J., Subramanian, S. v., & Carson, R. (2002). Zip Code Caveat: Bias Due to Spatiotemporal Mismatches Between Zip Codes and US Census–Defined Geographic Areas—The Public Health Disparities Geocoding Project. *American Journal of Public Health*, 92(7), 1100–1102. https://doi.org/10.2105/AJPH.92.7.1100

Krive, J., Patel, M., Gehm, L., Mackey, M., Kulstad, E., Li, J. J., Lussier, Y. A., & Boyd, A. D. (2015). The complexity and challenges of the international classification of diseases, ninth revision, clinical modification to international classification of diseases, 10th revision, clinical modification transition in EDs. *American Journal of Emergency Medicine*, 33(5), 713–718. https://doi.org/10.1016/j.ajem.2015.03.001

Kudamatsu, M., Persson, T., & Strömberg, D. (2012). Weather and Infant Mortality in Africa. Centre for Economic Policy Research, 1–77. http://www-2.iies.su.se/Nobel2012/Papers/Kudamatsu.pdf

Kulinkina, A. v., Mohan, V. R., Francis, M. R., Kattula, D., Sarkar, R., Plummer, J. D., Ward, H., Kang, G., Balraj, V., & Naumova, E. N. (2016). Seasonality of water quality and diarrheal disease counts in urban and rural settings in south India. *Scientific Reports*, 6(1), 1–12. https://doi.org/10.1038/srep20521

Kusiak, A., Verma, A., & Wei, X. (2013). A data-mining approach to predict influent quality. *Environmental Monitoring and Assessment*, 185(3), 2197–2210. https://doi.org/10.1007/s10661-012-2701-2

Lake, I. R., Gillespie, I. A., Bentham, G., Nichols, G. L., Lane, C., Adak, G. K., & Threlfall, E. J. (2009). A re-evaluation of the impact of temperature and climate change on foodborne illness. *Epidemiology and Infection*, 137(11), 1538–1547. https://doi.org/10.1017/S0950268809002477

Lake, I. R., Nichols, G., Bentham, G., Harrison, F. C. D., Hunter, P. R., & Kovats, R. S. (2007). Cryptosporidiosis Decline after Regulation, England and Wales, 1989–2005. *Emerging Infectious Diseases,* 13(4), 623–625. https://doi.org/10.3201/eid1304.060890

Lal, A., Hales, S., Kirk, M., Baker, M. G., & French, N. P. (2016). Spatial and temporal variation in the association between temperature and salmonellosis in NZ. *Australian and New Zealand Journal of Public Health,* 40(2), 165–169. https://doi.org/10.1111/1753-6405.12413

Lall, U., AghaKouchak, A., Arumugam, S., Brown, C., Mccabe, G., & Pulwarty, R. (2018). Water. In D. R. Reidmiller, C. W. Avery, D. R. Easterling, K. E. Kunkel, K. L. M. Lewis, T. K. Maycock, & B. C. Stewart (Eds.), *Impacts, Risks, and Adaptation in the United States: Fourth National Climate Assessment, Volume II* (pp. 145–173). U.S. Global Change Research Program. https://doi.org/10.7930/NCA4.2018.CH3

Lama, J. R., Seas, C. R., León-Barúa, R., Gotuzzo, E., & Sack, R. B. (2004). Environmental temperature, cholera, and acute diarrhoea in adults in Lima, Peru. *Journal of Health, Population, and Nutrition,* 22(4), 399–403. http://www.ncbi.nlm.nih.gov/pubmed/15663172

Lanata, C. F., Fischer-Walker, C. L., Olascoaga, A. C., Torres, C. X., Aryee, M. J., & Black, R. E. (2013). Global Causes of Diarrheal Disease Mortality in Children <5 Years of Age: A Systematic Review. *PLoS ONE*, 8(9), e72788. https://doi.org/10.1371/journal.pone.0072788

Langbein, W. G. (1967). Hydroclimate. In R. W. Fallbridge (Ed.), *Encyclopedia of Atmospheric Sciences and Astrogeology* (pp. 447–451). Reinhold.

Leclerc, H., Schwartzbrod, L., & Dei-Cas, E. (2002). Microbial agents associated with waterborne diseases. *Critical Reviews in Microbiology*, 28(4), 371–409. https://doi.org/10.1080/1040-840291046768

Lee, D., Chang, H. H., Sarnat, S. E., & Levy, K. (2019). precipitation and salmonellosis incidence in Georgia, USA: Interactions between extreme rainfall events and antecedent rainfall conditions. *Environmental Health Perspectives*, 127(9), 097005. https://doi.org/10.1289/EHP4621

Lee, H., Lau, S.-L., Kayhanian, M., & Stenstrom, M. K. (2004). Seasonal first flush phenomenon of urban stormwater discharges. *Water Research*, 38(19), 4153–4163. https://doi.org/10.1016/j.watres.2004.07.012

Lee, H. S., Her, M., Levine, M., & Moore, G. E. (2013). Time series analysis of human and bovine brucellosis in South Korea from 2005 to 2010. *Preventive Veterinary Medicine*, 110(2), 190–197. https://doi.org/10.1016/j.prevetmed.2012.12.003

Lee, H. S., Ha Hoang, T. T., Pham-Duc, P., Lee, M., Grace, D., Phung, D. C., Thuc, V. M., & Nguyen-Viet, H. (2017). Seasonal and geographical distribution of bacillary dysentery (shigellosis) and associated climate risk factors in Kon Tam Province in Vietnam from 1999 to 2013. *Infectious Diseases of Poverty*, 6(1). https://doi.org/10.1186/s40249-017-0325-z

Lee, R. M., Lessler, J., Lee, R. A., Rudolph, K. E., Reich, N. G., Perl, T. M., & Cummings, D. A. T. (2013). Incubation periods of viral gastroenteritis: A systematic review. *BMC Infectious Diseases*, 13(1), 1–11. https://doi.org/10.1186/1471-2334-13-446

Levy, K., Hubbard, A. E., & Eisenberg, J. N. S. (2009). Seasonality of rotavirus disease in the tropics: a systematic review and meta-analysis. *International Journal of Epidemiology*, 38(6), 1487–1496. https://doi.org/10.1093/ije/dyn260

Levy, K., Hubbard, A. E., Nelson, K. L., & Eisenberg, J. N. S. (2009). Drivers of water quality variability in northern coastal Ecuador. *Environmental Science & Technology*, 43(6), 1788–1797. https://doi.org/10.1021/es8022545

Levy, K., Woster, A. P., Goldstein, R. S., & Carlton, E. J. (2016). Untangling the impacts of climate change on waterborne diseases: A systematic review of relationships between diarrheal diseases and temperature, rainfall, flooding, and drought. *Environmental Science & Technology*, 50(10), 4905–4922. https://doi.org/10.1021/acs.est.5b06186

Levy, M. C., Collender, P. A., Carlton, E. J., Chang, H. H., Strickland, M. J., Eisenberg, J. N. S., & Remais, J. V. (2019). Spatiotemporal Error in Rainfall Data: Consequences for Epidemiologic Analysis of Waterborne Diseases. *American Journal of Epidemiology*, 188(5), 950–959. https://doi.org/10.1093/aje/kwz010

Leyk, S., McCormick, B. J. J., & Nuckols, J. R. (2011). Effects of varying temporal scale on spatial models of mortality patterns attributed to pediatric diarrhea. *Spatial and Spatio-Temporal Epidemiology*, 2(2), 91–101. https://doi.org/10.1016/j.sste.2011.03.002

Leyk, S., Phillips, T. P., Smith, J. M., & Nuckols, J. R. (2011). Spatio-temporal trends of diarrheal mortality of children in association with hydrographic regions of Brazil. *Cartography and Geographic Information Science,* 38(2), 223–232. https://doi.org/10.1559/15230406382223

Li, Z., Wang, L., Sun, W., Hou, X., Yang, H., Sun, L., Xu, S., Sun, Q., Zhang, J., Song, H., & Lin, H. (2013). Identifying high-risk areas of bacillary dysentery and associated meteorological factors in Wuhan, China. *Scientific Reports*, 3(1), 3239. https://doi.org/10.1038/srep03239

Li, T., Yang, Z., & Wang, M. (2014). Temperature and atmospheric pressure may be considered as predictors for the occurrence of bacillary dysentery in Guangzhou, Southern China. *Revista Da Sociedade Brasileira de Medicina Tropical*, 47(3), 382–384. https://doi.org/10.1590/0037-8682-0144-2013

Li, K., Zhao, K., Shi, L., Wen, L., Yang, H., Cheng, J., Wang, X., & Su, H. (2016). Daily temperature change in relation to the risk of childhood bacillary dysentery among different age groups and sexes in a temperate city in China. *Public Health,* 131, 20–26. https://doi.org/10.1016/j.puhe.2015.10.011

Li, X., Wang, N., Ding, G., Li, X., & Xue, X. (2019). The relationship between meteorological factors and the risk of bacillary dysentery in Hunan Province, China. *Weather,* 74(4), 148–153. https://doi.org/10.1002/wea.3085

Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika,* 73(1), 13–22. https://doi.org/10.1093/biomet/73.1.13

Liang, M., Ding, X., Wu, Y., & Sun, Y. (2021). Temperature and risk of infectious diarrhea: A systematic review and meta-analysis. *Environmental Science and Pollution Research*, 1–11. https://doi.org/10.1007/S11356-021-15395-Z/TABLES/3

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, 2(3), 18–22.

Lin, C. J., Wade, T., & Hilborn, E. (2015). Flooding and Clostridium difficile Infection: A Case-Crossover Analysis. *International Journal of Environmental Research and Public Health*, 12(6), 6948–6964. https://doi.org/10.3390/ijerph120606948

Lin, S., Sun, M., Fitzgerald, E., & Hwang, S.-A. (2016). Did summer weather factors affect gastrointestinal infection hospitalizations in New York State? *Science of the Total Environment*, 550, 38–44. https://doi.org/10.1016/j.scitotenv.2015.12.153

Lippmann, S. J., Fuhrmann, C. M., Waller, A. E., & Richardson, D. B. (2013). Ambient temperature and emergency department visits for heat-related illness in North Carolina, 2007–2008. *Environmental Research*, 124, 35–42. https://doi.org/10.1016/j.envres.2013.03.009

Liu, Z., Liu, Y., Zhang, Y., Lao, J., Zhang, J., Wang, H., & Jiang, B. (2019). Effect of ambient temperature and its effect modifiers on bacillary dysentery in Jinan, China. *Science of the Total Environment*, 650, 2980–2986. https://doi.org/10.1016/j.scitotenv.2018.10.053

Liu, Z., Tong, M. X., Xiang, J., Dear, K., Wang, C., Ma, W., Lu, L., Liu, Q., Jiang, B., & Bi, P. (2020). Daily temperature and bacillary dysentery: Estimated effects, attributable risks, and future disease burden in 316 Chinese cities. *Environmental Health Perspectives*, 128(5), 1–9. https://doi.org/10.1289/EHP5779

Lloyd, S. J., Kovats, R. S., & Armstrong, B. G. (2007). Global diarrhoea morbidity, weather and climate. *Climate Research*, 34(2), 119–127. https://doi.org/10.3354/cr034119

Lo Iacono, G., Armstrong, B., Fleming, L. E., Elson, R., Kovats, S., Vardoulakis, S., & Nichols, G. L. (2017). Challenges in developing methods for quantifying the effects of weather and climate on water-associated diseases: A systematic review. *PLOS Neglected Tropical Diseases*, 11(6), e0005659. https://doi.org/10.1371/journal.pntd.0005659

Loehman, R. A., Elias, J., Douglass, R. J., Kuenzi, A. J., Mills, J. N., & Wagoner, K. (2012). Prediction of Peromyscus maniculatus (deer mouse) population dynamics in Montana, USA, using satellite-driven vegetation productivity and weather data. *Journal of Wildlife Diseases*, 48(2), 348–360. https://doi.org/10.7589/0090-3558-48.2.348

Lopman, B. A., Adak, G. K., Reacher, M., & Brown, D. W. G. (2003). Two Epidemiologic Patterns of Norovirus Outbreaks: Surveillance in England and Wales, 1992–2000. *Emerging Infectious Diseases*, 9(1), 71–77. https://doi.org/10.3201/eid0901.020175

Lopman, B. A., Armstrong, B., Atchison, C., & Gray, J. J. (2009). Host, weather and virological factors drive norovirus epidemiology: Time-series analysis of laboratory surveillance data in England and Wales. *PLoS ONE*, 4(8), e6671. https://doi.org/10.1371/journal.pone.0006671

Lopman, B. A., Hall, A. J., Curns, A. T., & Parashar, U. D. (2011). Increasing rates of gastroenteritis hospital discharges in US adults and the contribution of norovirus, 1996-2007. *Clinical Infectious Diseases*, 52(4), 466–474. https://doi.org/10.1093/cid/ciq163

Lowen, A. C., & Steel, J. (2014). Roles of Humidity and Temperature in Shaping Influenza Seasonality. *Journal of Virology*, 88(14), 7692–7695. https://doi.org/10.1128/JVI.03544-13

Lu, Y., & Zeger, S. L. (2007). On the equivalence of case-crossover and time series methods in environmental epidemiology. B*iostatistic*s, 8(2), 337–344. https://doi.org/10.1093/biostatistics/kxl013

Lumley, T., & Levy, D. (2000). Bias in the case-crossover design: Implications for studies of air pollution. *Environmetrics*, 11(6), 689–704. https://doi.org/10.1002/1099-095X(200011/12)11:6<689::AID-ENV439>3.0.CO;2-N

Luque Fernández, M. Á., Bauernfeind, A., Jiménez, J. D., Gil, C. L., Omeiri, N. el, & Guibert, D. H. (2009). Influence of temperature and rainfall on the evolution of cholera epidemics in Lusaka, Zambia, 2003–2006: analysis of a time series. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 103(2), 137–143. https://doi.org/10.1016/j.trstmh.2008.07.017

Ly, S., Charles, C., & Degré, A. (2011). Geostatistical interpolation of daily rainfall at catchment scale: the use of several variogram models in the Ourthe and Ambleve catchments, Belgium. *Hydrology and Earth System Sciences*, 15(7), 2259–2274. https://doi.org/10.5194/hess-15-2259-2011

MacDonald Gibson, J., & Pieper, K. J. (2017). Strategies to Improve Private-Well Water Quality: A North Carolina Perspective. *Environmental Health Perspectives*, 125(7), 076001. https://doi.org/10.1289/EHP890

Mackun, P., Wilson, S., Fischetti, T. R., & Goworowska, J. (2011). Population Distribution and Change: 2000 to 2010. http://www.census.gov/prod/cen2010/briefs/c2010br-01.pdf

Maclure, M. (1991). The case-crossover design: A method for studying transient effects on the risk of acute events. *American Journal of Epidemiology*, 133(2), 144–153. https://doi.org/10.1093/oxfordjournals.aje.a115853

Maclure, M., & Mittleman, M. A. (2000). Should we use a case-crossover design? *Annual Review of Public Health*, 21(1), 193–221. https://doi.org/10.1146/annurev.publhealth.21.1.193

MacNeil, N. (2015). The influence of slavery on the locations of North Carolina's industrial hog operations. 17th Annual North Carolina Environmental Justice Network Summit: The Role of International Solidarity in Achieving Environmental Justice.

Mallin, M. A., Williams, K. E., Esham, E. C., & Lowe, R. P. (2000). Effect of human development on bacteriological water quality in coastal watersheds. Ecological Applications, 10(4), 1047–1056. https://doi.org/10.1890/1051-0761(2000)010[1047:EOHDOB]2.0.CO;2

Manley, D. (2014). Scale, aggregation, and the modifiable areal unit problem. In M. M. Fischer & P. Nijkamp (Eds.), The Handbook of Regional Science (pp. 1157–1171). Springer.

Manson, S., Schroeder, J., Riper, D. v, Kugler, T., & Ruggles, S. (2021). IPUMS National Historical Geographic Information System: Version 16.0 [dataset].

Marsh, B., Parnell, A. M., & Joyner, A. M. (2010). Institutionalization of racial inequality in local political geographies. Urban Geography, 31(5), 691–709. https://doi.org/10.2747/0272-3638.31.5.691

Marshall, M. M., Naumovitz, D., Ortega, Y., & Sterling, C. R. (1997). Waterborne protozoan pathogens. Clinical Microbiology Reviews, 10(1), 67–85. https://doi.org/10.1128/CMR.10.1.67

Martinez-Urtaza, J., Huapaya, B., Gavilan, R. G., Blanco-Abad, V., Ansede-Bermejo, J., Cadarso-Suarez, C., Figueiras, A., & Trinanes, J. (2008). Emergence of asiatic vibrio diseases in south america in phase with El Niño. Epidemiology, 19(6), 829–837. https://doi.org/10.1097/EDE.0b013e3181883d43

Matsuda, F., Ishimura, S., Wagatsuma, Y., Higashi, T., Hayashi, T., Faruque, A. S. G., Sack, D. A., & Nishibuchi, M. (2008). Prediction of epidemic cholera due to Vibrio cholerae O1 in children younger than 10 years using climate data in Bangladesh. Epidemiology and Infection, 136(1), 73–79. https://doi.org/10.1017/S0950268807008175

Mattioli, M. C., Benedict, K. M., Murphy, J., Kahler, A., Kline, K. E., Longenberger, A., Mitchell, P. K., Watkins, S., Berger, P., Shanks, O. C., Barrett, C. E., Barclay, L., Hall, A. J., Hill, V., & Weltman, A. (2021). Identifying septic pollution exposure routes during a waterborne norovirus outbreak - A new application for human-associated microbial source tracking qPCR. Journal of Microbiological Methods, 180, 106091. https://doi.org/https://doi.org/10.1016/j.mimet.2020.106091

Maupin, M. A., Kenny, J. F., Hutson, S. S., Lovelace, J. K., Barber, N. L., & Linsey, K. S. (2014). Estimated Use of Water in the United States in 2010. Circular 1405. https://doi.org/10.3133/cir1405

McCullagh, P., & Nelder, J. A. (1989). *Generalized Linear Models (2nd ed.).* Chapman and Hall.

McIver, L. J., Imai, C., Buettner, P. G., Gager, P., Chan, V. S., Hashizume, M., Iddings, S. N., Kol, H., Raingsey, P. P., & Lyne, K. (2016). Diarrheal diseases and climate change in Cambodia. Asia Pacific Journal of Public Health, 28(7), 576–585. https://doi.org/10.1177/1010539516660190

McMichael, A. J., & Lindgren, E. (2011). Climate change: Present and future risks to health, and necessary responses. Journal of Internal Medicine, 270(5), 401–413. https://doi.org/10.1111/j.1365-2796.2011.02415.x

McMichael, A. J., Woodruff, R. E., & Hales, S. (2006). Climate change and human health: present and future risks. The Lancet, 367(9513), 859–869. https://doi.org/http://dx.doi.org/10.1016/S0140-6736(06)68079-3

Mead, P. S., Slutsker, L., Dietz, V., McCaig, L. F., Bresee, J. S., Shapiro, C., Griffin, P. M., & Tauxe, R. v. (1999). Food-related illness and death in the United States. Emerging Infectious Diseases, 5(5), 607–625. https://doi.org/10.3201/eid0505.990502

Mellor, J. E., Levy, K., Zimmerman, J., Elliott, M., Bartram, J., Carlton, E., Clasen, T., Dillingham, R., Eisenberg, J., Guerrant, R., Lantagne, D., Mihelcic, J., & Nelson, K. (2016). Planning for climate change: The need for mechanistic systems-based approaches to study climate change impacts on diarrheal diseases. *Science of the Total Environment*, 548–549, 82–90. https://doi.org/10.1016/j.scitotenv.2015.12.087

Mendelsohn, J., & Dawson, T. (2008). Climate and cholera in KwaZulu-Natal, South Africa: The role of environmental factors and implications for epidemic preparedness. International Journal of Hygiene and Environmental Health, 211(1–2), 156–162. https://doi.org/10.1016/j.ijheh.2006.12.002

Mertens, A., Balakrishnan, K., Ramaswamy, P., Rajkumar, P., Ramaprabha, P., Durairaj, N., Hubbard, A. E., Khush, R., Colford, J. M., & Arnold, B. F. (2019). Associations between high temperature, heavy rainfall, and diarrhea among young children in Rural Tamil Nadu, India: A prospective cohort study. *Environmental Health Perspectives*, 127(4), 047004. https://doi.org/10.1289/EHP3711

Messner, M., Shaw, S., Regli, S., Rotert, K., Blank, V., & Soller, J. (2006). An approach for developing a national estimate of waterborne disease due to drinking water and a national estimate model application. *Journal of Water and Health*, 04(Suppl 2), 201. https://doi.org/10.2166/wh.2006.024

Milazzo, A., Giles, L. C., Zhang, Y., Koehler, A. P., Hiller, J. E., & Bi, P. (2016). The effect of temperature on different Salmonella serotypes during warm seasons in a Mediterranean climate city, Adelaide, Australia. *Epidemiology and Infection*, 144(6), 1231–1240. https://doi.org/10.1017/S0950268815002587

Milazzo, A., Giles, L. C., Zhang, Y., Koehler, A. P., Hiller, J. E., & Bi, P. (2017). The effects of ambient temperature and heatwaves on daily Campylobacter cases in Adelaide, Australia, 1990–2012. *Epidemiology and Infection*, 145(12), 2603–2610. https://doi.org/10.1017/S095026881700139X

Milborrow, S. (2011). Derived from mda:mars by T. Hastie and R. Tibshirani. earth: Multivariate Adaptive Regression Splines.

Milborrow, S. (2020). Package "earth."

Miller, A. G., Ebelt, S., & Levy, K. (2022). Combined sewer overflows and gastrointestinal illness in Atlanta, 2002–2013: Evaluating the impact of infrastructure improvements. *Environmental Health Perspectives*, 130(5), 2002–2013. https://doi.org/10.1289/EHP10399

Mitchell, T. D., & Jones, P. D. (2005). An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *International Journal of Climatology*, 25(6), 693–712. https://doi.org/10.1002/joc.1181

Mittleman, M. A., Maclure, M., & Robins, J. M. (1995). Control sampling strategies for case-crossover studies: an assessment of relative efficiency. *American Journal of Epidemiology*, 142(1), 91–98. https://doi.org/10.1093/oxfordjournals.aje.a117550

Mittleman, M. A., & Mostofsky, E. (2014). Exchangeability in the case-crossover design. *International Journal of Epidemiology*, 43(5), 1645–1655. https://doi.org/10.1093/ije/dyu081

Moors, E., Singh, T., Siderius, C., Balakrishnan, S., & Mishra, A. (2013). Climate change and waterborne diarrhoea in northern India: Impacts and adaptation strategies. *Science of the Total Environment*, 468–469, S139–S151. https://doi.org/10.1016/j.scitotenv.2013.07.021

Morral-Puigmal, C., Martínez-Solanas, È., Villanueva, C. M., & Basagaña, X. (2018). Weather and gastrointestinal disease in Spain: A retrospective time series regression study. Environment International, 121(May), 649–657. https://doi.org/10.1016/j.envint.2018.10.003

Mourtzinis, S., Rattalino Edreira, J. I., Conley, S. P., & Grassini, P. (2017). From grid to field: Assessing quality of gridded weather data for agricultural applications. *European Journal of Agronomy*, 82, 163–172. https://doi.org/10.1016/j.eja.2016.10.013

Muggeo, V. M., & Hajat, S. (2009). Modelling the non-linear multiple-lag effects of ambient temperature on mortality in Santiago and Palermo: A constrained segmented distributed lag approach. *Occupational and Environmental Medicine*, 66(9), 584–591. https://doi.org/10.1136/oem.2007.038653

Mukabutera, A., Thomson, D., Murray, M., Basinga, P., Nyirazinyoye, L., Atwood, S., Savage, K. P., Ngirimana, A., & Hedt-Gauthier, B. L. (2016). Rainfall variation and child health: Effect of rainfall on diarrhea among under 5 children in Rwanda, 2010. *BMC Public Health*, 16(1). https://doi.org/10.1186/s12889-016-3435-9

Mukhopadhyay, A. K., Deb, A. K., Chowdhury, G., Debnath, F., Samanta, P., Saha, R. N., Manna, B., Bhattacharya, M. K., Datta, D., Okamoto, K., Bhadra, U. K., & Dutta, S. (2019). Post-monsoon waterlogging-associated upsurge of cholera cases in and around Kolkata metropolis, 2015. *Epidemiology and Infection*, 147, e167. https://doi.org/10.1017/S0950268819000529

Munnangi, S., & Boktor, S. W. (2018). Epidemiology, Study Design. In StatPearls [Internet]. StatPearls Publishing. https://www.ncbi.nlm.nih.gov/books/NBK470342/

Murphy, H. M., McGinnis, S., Blunt, R., Stokdyk, J., Wu, J., Cagle, A., Denno, D. M., Spencer, S., Firnstahl, A., & Borchardt, M. A. (2020). Septic systems and rainfall influence human fecal marker and indicator organism occurrence in private wells in Southeastern Pennsylvania. *Environmental Science & Technology*, 54(6), 3159–3168. https://doi.org/10.1021/acs.est.9b05405

Murray, A., Hall, A., Weaver, J., & Kremer, F. (2021). Methods for Estimating Locations of Housing Units Served by Private Domestic Wells in the United States Applied to 2010. *JAWRA Journal of the American Water Resources Association*, 57(5), 828–843. https://doi.org/10.1111/1752-1688.12937

Musengimana, G., Mukinda, F., Machekano, R., & Mahomed, H. (2016). Temperature variability and occurrence of diarrhoea in children under five-years-old in Cape Town metropolitan sub-districts. *International Journal of Environmental Research and Public Health*, 13(9), 859. https://doi.org/10.3390/ijerph13090859

Napolitano, A. (2016). Cyclostationarity: New trends and applications. *Signal Processing*, 120, 385–408. https://doi.org/10.1016/j.sigpro.2015.09.011

Nataro, J. P., & Kaper, J. B. (1998). Diarrheagenic *Escherichia coli*. *Clinical Microbiology Reviews*, 11(1), 142–201. https://doi.org/10.1128/CMR.11.1.142

Naumova, E. N., Chen, J. T., Griffiths, J. K., Matyas, B. T., Estessmargiassi, S. A., & Morris, R. D. (2000). Use of passive surveillance data to study temporal and spatial variation in the incidence of giardiasis and cryptosporidiosis. *Public Health Reports*, 115(5), 436–447. https://doi.org/10.1093/phr/115.5.436

Naumova, E. N., Christodouleas, J., Hunter, P. R., & Syed, Q. (2005). Effect of precipitation on seasonal variability in cryptosporidiosis recorded by the North West England surveillance system in 1990-1999. *Journal of Water and Health*, 3(2), 185–196.

Naumova, E. N., Jagai, J. S., Matyas, B., DeMaria, A., MacNeill, I. B., & Griffiths, J. K. (2007). Seasonality in six enterically transmitted diseases and ambient temperature. *Epidemiology and Infection*, 135(2), 281–292. https://doi.org/10.1017/S0950268806006698

Navidi, W. (1998). Bidirectional case-crossover designs for exposures with time trends. *Biometrics*, 54(2), 596–605. https://doi.org/10.2307/3109766

Navidi, W., & Weinhandl, E. (2002). Risk set sampling for case-crossover designs. *Epidemiology*, 13(1), 100–105. https://doi.org/10.1097/00001648-200201000-00016

NC Department of Health and Human Services (NC DHHS). (n.d.). EHS: On-Site Wastewater Treatment and Dispersal Systems Program Resources. 2022. Retrieved July 22, 2022, from https://ehs.dph.ncdhhs.gov/oswp/resources.htm

NC DHHS. (2021). Private Wells: Facts and Figures. NC Department of Health and Human Services (NC DHHS). https://epi.dph.ncdhhs.gov/oee/wellwater/figures.html

NCDEQ-DWR. (2020). Animal Feed Operation Permits [dataset]. North Carolina Department of Environmental Quality, Division of Water Resources. https://ncdenr.maps.arcgis.com/apps/webappviewer/index.html?id=85ae6392d0e94010a305eedf06e3f288

Ngo-Duc, T., Polcher, J., & Laval, K. (2005a). NCC Dataset.

Ngo-Duc, T., Polcher, J., & Laval, K. (2005b). A 53-year forcing data set for land surface models. *Journal of Geophysical Research: Atmospheres*, 110(D6), D06116 (1-13). https://doi.org/10.1029/2004JD005434

Nichols, G., Lane, C., Asgari, N., Verlander, N. Q., & Charlett, A. (2009). Rainfall and outbreaks of drinking water related disease and in England and Wales. *Journal of Water and Health*, 7(1), 1–8. https://doi.org/10.2166/wh.2009.143

Nitta, H., Yamazaki, S., Omori, T., & Sato, T. (2010). An introduction to epidemiologic and statistical methods useful in environmental epidemiology. *Journal of Epidemiology*, 20(3), 177–184. https://doi.org/10.2188/jea.JE20100010

Nordhaus, W. D. (2010). Economic aspects of global warming in a post-Copenhagen environment. *Proceedings of the National Academy of Sciences*, 107(26), 11721–11726. http://www.pnas.org/content/107/26/11721.short

Nordhaus, W. D., & Sztorc, P. (2013). DICE 2013R: Introduction and user's manual (2nd Ed.). http://www.econ.yale.edu/~nordhaus/homepage/documents/DICE_Manual_103113r2_000.pdf

North Carolina Institute of Medicine. (2018). North Carolina County Health Profiles [WWW Document]. http://nciom.org/map/

Norton, J. M., Wing, S., Lipscomb, H. J., Kaufman, J. S., Marshall, S. W., & Cravey, A. J. (2007). Race, wealth, and solid waste facilities in North Carolina. *Environmental Health Perspectives*, 115(9), 1344–1350. https://doi.org/10.1289/ehp.10161

Nygård, K., Andersson, Y., Røttingen, J. A., Svensson, Å., Lindbäck, J., Kistemann, T., & Giesecke, J. (2004). Association between environmental risk factors and campylobacter infections in Sweden. *Epidemiology and Infection*, 132(2), 317–325. https://doi.org/10.1017/S0950268803001900

Odoi, A., Martin, S. W., Michel, P., Holt, J., Middleton, D., & Wilson, J. (2004). Determinants of the geographical distribution of endemic giardiasis in Ontario, Canada: a spatial modelling approach. *Epidemiology and Infection*, 132(5), 967–976.

Odoi, A., Martin, S. W., Michel, P., Middleton, D., Holt, J., & Wilson, J. (2004). Investigation of clusters of giardiasis using GIS and a spatial scan statistic. International Journal of Health Geographics, 3(1), 11. http://ij-healthgeographics.biomedcentral.com/articles/10.1186/1476-072X-3-11

Onozuka, D., & Hashizume, M. (2011). Weather variability and paediatric infectious gastroenteritis. *Epidemiology and Infection*, 139(09), 1369–1378.

Onozuka, D., Hashizume, M., & Hagihara, A. (2010). Effects of weather variability on infectious gastroenteritis. *Epidemiology and Infection*, 138(2), 236–243. https://doi.org/10.1017/S0950268809990574

Openshaw, S. (1978). An Empirical Study of Some Zone-Design Criteria. Environment and Planning A: Economy and Space, 10(7), 781–794. https://doi.org/10.1068/a100781

Openshaw, S. (1984a). Ecological fallacies and the analysis of areal census data. Environment and Planning A, 16(1), 17–31. http://journals.sagepub.com/doi/10.1068/a160017

Openshaw, S. (1984b). The modifiable areal unit problem: Concepts and techniques in modern geography. Geo Books. https://www.uio.no/studier/emner/sv/iss/SGO9010/openshaw1983.pdf

Openshaw, S., & Taylor, P. J. (1979). A million or so correlation coefficients: Three experiments on the modifiable areal unit problem. In N. Wrigley (Ed.), *Statistical Applicaions in the Spatial Sciences* (pp. 127–144). Pion. http://www.csiss.org/GISPopSci/workshops/2009/UCSB/readings/Openshaw-Taylor-1979.pdf

Ordiz, M. I., Shaikh, N., Trehan, I., Maleta, K., Stauber, J., Shulman, R., Devaraj, S., Tarr, P. I., & Manary, M. J. (2016). Environmental enteric dysfunction is associated with poor linear growth and can be identified by host fecal mRNAs. *Journal of Pediatric Gastroenterology and Nutrition*, 63(5), 453–459. https://doi.org/10.1097/MPG.0000000000001315

Orlowsky, B., & Seneviratne, S. I. (2012). Global changes in extreme events: regional and seasonal dimension. *Climatic Change*, 110(3–4), 669–696. https://doi.org/10.1007/s10584-011-0122-9

Parkes, M. W., Morrison, K. E., Bunch, M. J., & Venema, H. D. (2008). Ecohealth and watersheds: Ecosystem approaches to re-integrate water resources management with health and well-being. In *Network for Ecosystem Sustainability and Health (Publication Series No. 2) and the International Institute for Sustainable Development*. https://www.iisd.org/publications/report/ecohealth-and-watersheds-ecosystem-approaches-re-integrate-water-resources

Pascual, M., Cazelles, B., Bouma, M. J., Chaves, L. F., & Koelle, K. (2008). Shifting patterns: malaria dynamics and rainfall variability in an African highland. *Proceedings of the Royal Society B-Biological Sciences*, 275(1631), 123–132.

Pascual, M., Chaves, L. F., Cash, B. A., Rodó, X., & Yunus, M. (2008). Predicting endemic cholera: the role of climate variability and disease dynamics. *Climate Research*, 36, 131–140.

Pascual, M., Rodó, X., Ellner, S. P., Colwell, R. R., & Bouma, M. J. (2000). Cholera Dynamics and El Niño-Southern Oscillation. *Science*, 289(5485), 1766–1769. http://www.sciencemag.org/cgi/doi/10.1126/science.289.5485.1766

Pascual, M., Roy, M., & Laneri, K. (2011). Simple models for complex systems: exploiting the relationship between local and global densities. *Theoretical Ecology*, 4(2), 211–222. http://dx.doi.org/10.1007/s12080-011-0116-2

Patz, J. A., Engelberg, D., & Last, J. (2000). The effects of changing weather on public health. *Annual Review of Public Health*, 21(1), 271–307.

Patz, J. A., Frumkin, H., Holloway, T., Vimont, D. J., & Haines, A. (2014). Climate change: Challenges and opportunities for global health. *JAMA*, 53726. https://doi.org/10.1001/jama.2014.13186

Payment, P., Richardson, L., Siemiatycki, J., Dewar, R., Edwardes, M., & Franco, E. (1991). A randomized trial to evaluate the risk of gastrointestinal disease due to consumption of drinking water meeting current microbiological standards. *American Journal of Public Health*, 81(6), 703–708. https://doi.org/10.2105/AJPH.81.6.703

Paz, S. (2009). Impact of temperature variability on cholera incidence in Southeastern Africa, 1971–2006. *EcoHealth*, 6(3), 340–345. https://doi.org/10.1007/s10393-009-0264-7

Penakalapati, G., Swarthout, J., Delahoy, M. J., McAliley, L., Wodnik, B., Levy, K., & Freeman, M. C. (2017). Exposure to animal feces and human health: A systematic review and proposed research priorities. *Environmental Science and Technology*, 51(20), 11537–11552. https://doi.org/10.1021/acs.est.7b02811

Peng, R. D., & Dominici, F. (2008). *Statistical Methods for Environmental Epidemiology with R*. Springer New York. https://doi.org/10.1007/978-0-387-78167-9

Peng, R. D., Dominici, F., & Louis, T. A. (2006). Model choice in time series studies of air pollution and mortality. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169(2), 179–203. https://doi.org/10.1111/j.1467-985X.2006.00410.x

Perrakis, K., Gryparis, A., Schwartz, J., Tertre, A. Le, Katsouyanni, K., Forastiere, F., Stafoggia, M., & Samoli, E. (2014). Controlling for seasonal patterns and time varying confounders in time-series epidemiological models: A simulation study. *Statistics in Medicine*, 33(28), 4904–4918. https://doi.org/10.1002/sim.6271

Petersen, T., & Koput, K. W. (1992). Time-Aggregation Bias in Hazard-Rate Models with Covariates. Sociological Methods & Research, 21(1), 25–51. https://doi.org/10.1177/0049124192021001002

Philipsborn, R., Ahmed, S. M., Brosi, B. J., & Levy, K. (2016). Climatic drivers of diarrheagenic Escherichia coli incidence: A systematic review and meta-analysis. Journal of Infectious Diseases, 214(1), 6–15. https://doi.org/10.1093/infdis/jiw081

Phung, D., Chu, C., Rutherford, S., Nguyen, H. L. T., Luong, M. A., Do, C. M., & Huang, C. (2017). Heavy rainfall and risk of infectious intestinal diseases in the most populous city in Vietnam. Science of the Total Environment, 580, 805–812. https://doi.org/10.1016/j.scitotenv.2016.12.027

Phung, D., Huang, C., Rutherford, S., Chu, C., Wang, X., Nguyen, M., Nguyen, N. H., Manh, C. D., & Nguyen, T. H. (2015). Association between climate factors and diarrhoea in a Mekong Delta area. International Journal of Biometeorology, 59(9), 1321–1331. https://doi.org/10.1007/s00484-014-0942-1

Phung, D., Rutherford, S., Chu, C., Wang, X., Nguyen, M., Nguyen, N. H., Manh Do, C., Nguyen, T. H., & Huang, C. (2015). Temperature as a risk factor for hospitalisations among young children in the Mekong Delta area, Vietnam. Occupational and Environmental Medicine, 72(7), 529–535. https://doi.org/10.1136/oemed-2014-102629

Piantadosi, S., Byar, D. P., & Green, S. B. (1988). The ecological fallacy. American Journal of Epidemiology, 127(5), 893–904. https://doi.org/10.1093/oxfordjournals.aje.a114892

Pires, S. M., Fischer-Walker, C. L., Lanata, C. F., Devleesschauwer, B., Hall, A. J., Kirk, M. D., Duarte, A. S. R., Black, R. E., & Angulo, F. J. (2015). Aetiology-specific estimates of the global and regional incidence and mortality of diarrhoeal diseases commonly transmitted through food. PLoS ONE, 10(12). https://doi.org/10.1371/journal.pone.0142927

Poulsen, M. N., Pollak, J., Sills, D. L., Casey, J. A., Rasmussen, S. G., Nachman, K. E., Cosgrove, S. E., Stewart, D., & Schwartz, B. S. (2018). Residential proximity to high-density poultry operations associated with campylobacteriosis and infectious diarrhea. International Journal of Hygiene and Environmental Health, 221(2), 323–333. https://doi.org/10.1016/j.ijheh.2017.12.005

Prasad, A. M., Iverson, L. R., & Liaw, A. (2006). newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems*, 9(2), 181–199. https://doi.org/10.1007/s10021-005-0054-1

Prein, A. F., Rasmussen, R. M., Ikeda, K., Liu, C., Clark, M. P., & Holland, G. J. (2017). The future intensification of hourly precipitation extremes. *Nature Climate Change*, 7(1), 48–52. https://doi.org/10.1038/nclimate3168

PRISM Climate Group. (2004). Parameter-elevation regressions on independent slopes model (PRISM). Oregon State University. http://www.prism.oregonstate.edu/

PRISM Climate Group. (2016). Descriptions of PRISM spatial Climate Datasets for the Conterminous United States (Issue August). http://www.prism.oregonstate.edu/documents/PRISM_datasets.pdf

Prüss-Ustün, A., Bartram, J., Clasen, T., Colford, J. M., Cumming, O., Curtis, V., Bonjour, S., Dangour, A. D., de France, J., Fewtrell, L., Freeman, M. C., Gordon, B., Hunter, P. R., Johnston, R. B., Mathers, C., Mäusezahl, D., Medlicott, K., Neira, M., Stocks, M., …

Cairncross, S. (2014). Burden of disease from inadequate water, sanitation and hygiene in low- and middle-income settings: A retrospective analysis of data from 145 countries. *Tropical Medicine and International Health*, 19(8), 894–905. https://doi.org/10.1111/tmi.12329

Quist, A. J. L., Fliss, M. D., Wade, T. J., Delamater, P. L., Richardson, D. B., & Engel, L. S. (2022). Hurricane flooding and acute gastrointestinal illness in North Carolina. *Science of the Total Environment*, 809, 151108. https://doi.org/10.1016/j.scitotenv.2021.151108

Quist, A. J. L., Holcomb, D. A., Fliss, M. D., Delamater, P. L., Richardson, D. B., & Engel, L. S. (2022). Exposure to industrial hog operations and gastrointestinal illness in North Carolina, USA. *Science of the Total Environment*, 830, 154823. https://doi.org/10.1016/j.scitotenv.2022.154823

Raisz, E. (1940). Landforms of the United States. In W. W. Atwood (Ed.), *The physiographic provinces of North America*. Blaisdell Publishing Co. https://www.ncpedia.org/our-state-geography-snap-three

Ravindra, K., Rattan, P., Mor, S., & Aggarwal, A. N. (2019). Generalized additive models: Building evidence of air pollution, climate change and human health. *Environment International*, 132, 104987. https://doi.org/10.1016/j.envint.2019.104987

Reacher, M., McKenzie, K., Lane, C., Nichols, T., Kedge, I., Iversen, A., Hepple, P., Walter, T., Laxton, C., & Simpson, J. (2004). Health impacts of flooding in Lewes: a comparison of reported gastrointestinal and other illness and mental health in flooded and non-flooded households. *Communicable Disease and Public Health*, 7(1), 39–46. https://pubmed.ncbi.nlm.nih.gov/15137280/

Redman, R. L., Nenn, C. A., Eastwood, D., & Gorelick, M. H. (2007). Pediatric emergency department visits for diarrheal illness increased after release of undertreated sewage. *Pediatrics*, 120(6), e1472–e1475. https://doi.org/10.1542/peds.2007-0283

Reyburn, R., Kim, D. R., Emch, M., Khatib, A., von Seidlein, L., & Ali, M. (2011). Climate variability and the outbreaks of cholera in Zanzibar, East Africa: A time series analysis. *The American Journal of Tropical Medicine and Hygiene*, 84(6), 862–869. https://doi.org/10.4269/ajtmh.2011.10-0277

Roberts, S., & Martin, M. A. (2007). A distributed lag approach to fitting non-linear dose–response models in particulate matter air pollution time series investigations. *Environmental Research*, 104(2), 193–200. https://doi.org/10.1016/j.envres.2007.01.009

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C. J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J. K., Walker, J. P., Lohmann, D., & Toll, D. (2004). The Global Land Data Assimilation System. *Bulletin of the American Meteorological Society*, 85(3), 381–394. https://doi.org/10.1175/BAMS-85-3-381

Roderick, P., Wheeler, J., Cowdex, J., Sockett, P., Skinner, R., Mortimer, P., Rowe, B., & Rodriques, L. (1995). A pilot study of infectious intestinal disease in England. *Epidemiology and Infection*, 114(2), 277–288. https://doi.org/10.1017/S0950268800057940

Rodó, X., Pascual, M., Fuchs, G., & Faruque, A. S. G. (2002). ENSO and cholera: A nonstationary link related to climate change? *Proceedings of the National Academy of Sciences*, 99(20), 12901–12906. http://www.pnas.org/content/99/20/12901.short

Rohayem, J. (2009). Norovirus seasonality and the potential impact of climate change. *Clinical Microbiology and Infection,* 15(6), 524–527. https://doi.org/10.1111/j.1469-0691.2009.02846.x

Romero, R., Guijarro, J. A., Ramis, C., & Alonso, S. (1998). A 30-year (1964-1993) daily rainfall data base for the Spanish Mediterranean regions: First exploratory study. *International Journal of Climatology*, 18(5), 541–560. https://doi.org/10.1002/(sici)1097-0088(199804)18:5<541::aid-joc270>3.0.co;2-n

Rosenthal, J. (2009). Climate change and the geographic distribution of infectious diseases. EcoHealth, 6(4), 489–495.

Roy, S. L., Scallan, E., & Beach, M. J. (2006). The rate of acute gastrointestinal illness in developed countries. *Journal of Water and Health*, 4(S2), 31–69. https://doi.org/10.2166/wh.2006.017

Scallan, E., Griffin, P. M., Angulo, F. J., Tauxe, R. V., & Hoekstra, R. M. (2011). Foodborne illness acquired in the United States-Unspecified agents. *Emerging Infectious Diseases*, 17(1), 16–22. https://doi.org/10.3201/eid1701.P21101

Scallan, E., Griffin, P. M., McLean, H. Q., & Mahon, B. E. (2018). Hospitalisations due to bacterial gastroenteritis: A comparison of surveillance and hospital discharge data. *Epidemiology and Infection*, 146(8), 954–960. https://doi.org/10.1017/S0950268818000882

Scallan, E., Hoekstra, R. M., Angulo, F. J., Tauxe, R. V., Widdowson, M.-A., Roy, S. L., Jones, J. L., & Griffin, P. M. (2011). Foodborne illness acquired in the United States-Major pathogens. *Emerging Infectious Diseases*, 17(1), 7–15. https://doi.org/10.3201/eid1701.P11101

Scandura, J. E., & Sobsey, M. D. (1997). Viral and bacterial contamination of groundwater from on-site sewage treatment systems. *Water Science and Technology*, 35(11–12), 141–146. https://doi.org/10.2166/wst.1997.0724

Schär, C., Ban, N., Fischer, E. M., Rajczak, J., Schmidli, J., Frei, C., Giorgi, F., Karl, T. R., Kendon, E. J., Tank, A. M. G. K., O'Gorman, P. A., Sillmann, J., Zhang, X., & Zwiers, F. W. (2016). Percentile indices for assessing changes in heavy precipitation events. *Climatic Change*, 137(1–2), 201–216. https://doi.org/10.1007/s10584-016-1669-2

Schlenker, W., Hanemann, M., & Fisher, A. (2005). Will U.S. Agriculture Really Benefit from Global Warming? Accounting for Irrigation in the Hedonic Approach. *The American Economic Review*, 95(1), 395–406.

Schlenker, W., Hanemann, W. M., & Fisher, A. C. (2006). The impact of global warming on U.S. agriculture: An econometric analysis of optimal growing conditions. *Review of Economics and Statistics*, 88(1), 113–125. https://doi.org/10.1162/003465306775565684

Schlenker, W., Hanemann, W. M., & Fisher, A. C. (2007). Water Availability, Degree Days, and the Potential Impact of Climate Change on Irrigated Agriculture in California. *Climatic Change*, 81(1), 19–38. https://doi.org/10.1007/s10584-005-9008-z

Schlenker, W., & Roberts, M. J. (2006). Nonlinear Effects of Weather on Corn Yields. *Review of Agricultural Economics*, 28(3), 391–398. https://doi.org/10.1111/j.1467-9353.2006.00304.x

Schlenker, W., & Roberts, M. J. (2009). Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. *Proceedings of the National Academy of Sciences*, 106(37), 15594–15598. https://doi.org/10.1073/pnas.0906865106

Schmidt, W.-P. (2014). The elusive effect of water and sanitation on the global burden of disease. *Tropical Medicine & International Health*, 19(5), 522–527. https://doi.org/10.1111/tmi.12286

Schnitzler, J., Benzler, J., Altmann, D., Mücke, I., & Krause, G. (2007). Survey on the population's needs and the public health response during floods in Germany 2002. *Journal of Public Health Management and Practice*, 13(5), 461–464. https://doi.org/10.1097/01.PHH.0000285197.23932.3e

Schumacher, I. (2018). The aggregation dilemma in climate change policy evaluation. *Climate Change Economics*, 9(3), 1–21. https://doi.org/10.1142/S2010007818500082

Schwartz, J. (2000a). Assessing confounding, effect modification, and thresholds in the association between ambient particles and daily deaths. *Environmental Health Perspectives*, 108(6), 563–568. https://doi.org/10.1289/ehp.00108563

Schwartz, J. (2000b). The distributed lag between air pollution and daily deaths. *Epidemiology*, 11(3), 320–326. https://doi.org/10.1097/00001648-200005000-00016

Schwartz, J., Spix, C., Touloumi, G., Bachárová, L., Barumamdzadeh, T., Le Tertre, A., Piekarksi, T., Ponce De Leon, A., Pönkä, A., Rossi, G., Saez, M., & Schouten, J. P. (1996). Methodological issues in studies of air pollution and daily counts of deaths or hospital admissions. *Journal of Epidemiology and Community Health*, 50(SUPPL. 1), 3–4. https://doi.org/10.1136/jech.50.suppl_1.s3

Seidu, R., Stenström, T. A., & Löfman, O. (2013). A comparative cohort study of the effect of rainfall andtemperature on diarrhoeal disease in faecal sludge andnon-faecal sludge applying communities, Northern Ghana. *Journal of Water and Climate Change*, 4(2), 90–102. https://doi.org/10.2166/wcc.2013.032

Semenza, J. C., Herbst, S., Rechenburg, A., Suk, J. E., Höser, C., Schreiber, C., & Kistemann, T. (2012). Climate change impact assessment of food- and waterborne diseases. *Critical Reviews* *in Environmental Science and Technology*, 42(8), 857–890. https://doi.org/10.1080/10643389.2010.534706

Semenza, J. C., Höser, C., Herbst, S., Rechenburg, A., Suk, J. E., Frechen, T., & Kistemann, T. (2012). Knowledge mapping for climate change and food- and waterborne diseases. *Critical Reviews in Environmental Science and Technology,* 42(4), 378–411. https://doi.org/10.1080/10643389.2010.518520

Shafran-Nathan, R., Levy, I., Levin, N., & Broday, D. M. (2017). Ecological bias in environmental health studies: The problem of aggregation of multiple data sources. *Air Quality, Atmosphere & Health*, 10(4), 411–420. https://doi.org/10.1007/s11869-016-0436-x

Shortridge, J. E., & Guikema, S. D. (2014). Public health and pipe breaks in water distribution systems: Analysis with internet search volume as a proxy. *Water Research*, 53, 26–34. https://doi.org/10.1016/j.watres.2014.01.013

Shumway, D. S., & Stoffer, R. H. (2006). *Time Series Analysis and its Applications: with R Examples*. Springer New York. https://doi.org/10.1007/0-387-36276-2

Shumway, D. S., & Stoffer, R. H. (2017). *Time Series Analysis and its Applications: with R Examples (4th ed.)*. Springer. https://doi.org/10.1007/978-3-319-52452-8

Sillmann, J., Kharin, V. v, Zwiers, F. W., Zhang, X., & Bronaugh, D. (2013). Climate extremes indices in the CMIP5 multimodel ensemble: Part 2. Future climate projections. *Journal of Geophysical Research: Atmospheres*, 118(6), 2473–2493. https://doi.org/10.1002/jgrd.50188

Singh, R. B., Hales, S., de Wet, N., Raj, R., Hearnden, M., & Weinstein, P. (2001). The influence of climate variation and change on diarrheal disease in the Pacific Islands. *Environmental Health Perspectives*, 109(2), 155–159. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1240636

Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K., & et al. (2007). Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, 2007 [Internet].

Son, J.-Y., Muenich, R. L., Schaffer-Smith, D., Miranda, M. L., & Bell, M. L. (2021). Distribution of environmental justice metrics for exposure to CAFOs in North Carolina, USA. *Environmental Research*, 195(October 2020), 110862. https://doi.org/10.1016/j.envres.2021.110862

Soneja, S., Jiang, C., Fisher, J., Upperman, C. R., Mitchell, C., & Sapkota, A. (2016). Exposure to extreme heat and precipitation events associated with increased risk of hospitalization for asthma in Maryland, U.S.A. *Environmental Health: A Global Access Science Source*, 15(1), 1–7. https://doi.org/10.1186/s12940-016-0142-z

Soneja, S., Jiang, C., Upperman, C. R., Murtugudde, R., Mitchell, C. S., Blythe, D., Sapkota, A. R., & Sapkota, A. (2016). Extreme precipitation events and increased risk of campylobacteriosis in

Maryland, U.S.A. *Environmental Research*, 149, 216–221. https://doi.org/10.1016/j.envres.2016.05.021

Song, Y.-J., Cheong, H.-K., Ki, M., Shin, J.-Y., Hwang, S., Park, M., Ki, M., & Lim, J. (2018). The Epidemiological Influence of Climatic Factors on Shigellosis Incidence Rates in Korea. International Journal of Environmental Research and Public Health, 15(10), 2209. https://doi.org/10.3390/ijerph15102209

Sowah, R., Zhang, H., Radcliffe, D., Bauske, E., & Habteselassie, M. Y. (2014). Evaluating the influence of septic systems and watershed characteristics on stream faecal pollution in suburban watersheds in Georgia, USA. Journal of Applied Microbiology, 117(5), 1500–1512. https://doi.org/10.1111/jam.12614

Spangler, K. R., Weinberger, K. R., & Wellenius, G. A. (2019). Suitability of gridded climate datasets for use in environmental epidemiology. Journal of Exposure Science and Environmental Epidemiology, 29(6), 777–789. https://doi.org/10.1038/s41370-018-0105-2

State Climate Office of North Carolina. (n.d.). Overview. Retrieved August 3, 2017, from http://climate.ncsu.edu/climate/ncclimate.html

Steinberg, D., Colla, P., & Martin, K. (1999). MARS user guide. Salford Systems.

Stelmach, R., & Clasen, T. (2015). Household Water Quantity and Health: A Systematic Review. International Journal of Environmental Research and Public Health, 12(6), 5954–5974. https://doi.org/10.3390/ijerph120605954

Sterk, A., Schijven, J., de Nijs, T., & de Roda Husman, A. M. (2013). Direct and indirect effects of climate change on the risk of infection by water-transmitted pathogens. Environmental Science and Technology, 47(22), 12648–12660. https://doi.org/10.1021/es403549s

Stieb, D. M., Judek, S., & Burnett, R. T. (2002). Meta-analysis of time-series studies of air pollution and mortality: Effects of gases and particles and the influence of cause of death, age, and season. Journal of the Air & Waste Management Association, 52(4), 470–484.

Stingone, J. A., & Wing, S. (2011). Poultry Litter Incineration as a Source of Energy: Reviewing the Potential for Impacts on Environmental Health and Justice. NEW SOLUTIONS: A Journal of Environmental and Occupational Health Policy, 21(1), 27–42. https://doi.org/10.2190/NS.21.1.g

Sugg, M. M., Konrad, C. E., & Fuhrmann, C. M. (2016). Relationships between maximum temperature and heat-related illness across North Carolina, USA. International Journal of Biometeorology, 60(5), 663–675. https://doi.org/10.1007/s00484-015-1060-4

Tam, C. C., Rodriguez, L. C., O'Brien, S. J., & Hajat, S. (2006). Temperature dependence of reported Campylobacter infection in England, 1989-1999. Epidemiology and Infection, 134(1), 119–125. https://doi.org/10.1017/S0950268805004899

Teschke, K., Bellack, N., Shen, H., Atwater, J., Chu, R., Koehoorn, M., MacNab, Y. C., Schreier, H., & Isaac-Renton, J. L. (2010). Water and sewage systems, socio-demographics, and duration of residence associated with endemic intestinal infectious diseases: a cohort study. *BMC Public Health*, 10, 767. https://doi.org/10.1186/1471-2458-10-767

Therneau, T. M., & Atkinson, E. J. (1997). An introduction to recursive partitioning using the RPART routines. In Technical Report no. 61. https://www.mayo.edu/research/documents/biostat-61pdf/doc-10026699

Therneau, T. M., & Atkinson, E. J. (2019). An introduction to recursive partitioning using the RPART routines. In *Division of Biostatistics* 61. https://cran.r-project.org/web/packages/rpart/vignettes/longintro.pdf

Thindwa, D., Chipeta, M. G., Henrion, M. Y. R., & Gordon, M. A. (2019). Distinct climate influences on the risk of typhoid compared to invasive non-typhoid Salmonella disease in Blantyre, Malawi. *Scientific Reports*, 9(1), 20310. https://doi.org/10.1038/s41598-019-56688-1

Thomas, D. C. (2009). Statistical Methods in Environmental Epidemiology. Oxford University Press. http://ebookcentral.proquest.com/lib/unc/detail.action?docID=472393

Thomas, K. M., Charron, D. F., Waltner-Toews, D., Schuster, C., Maarouf, A. R., & Holt, J. D. (2006). A role of high impact weather events in waterborne disease outbreaks in Canada, 1975 – 2001. *International Journal of Environmental Health Research*, 16(3), 167–180. https://doi.org/10.1080/09603120600641326

Thomas, N., Ebelt, S. T., Newman, A. J., Scovronick, N., D'Souza, R. R., Moss, S. E., Warren, J. L., Strickland, M. J., Darrow, L. A., & Chang, H. H. (2021). Time-series analysis of daily ambient temperature and emergency department visits in five US cities with a comparison of exposure metrics derived from 1-km meteorology products. *Environmental Health: A Global Access Science Source*, 20(1), 1–10. https://doi.org/10.1186/s12940-021-00735-w

Thompson, C. N., Zelner, J. L., Nhu, T. D. H., Phan, M. V., Hoang Le, P., Nguyen Thanh, H., Vu Thuy, D., Minh Nguyen, N., Ha Manh, T., van Hoang Minh, T., Lu Lan, V., Nguyen Van Vinh, C., Tran Tinh, H., von Clemm, E., Storch, H., Thwaites, G., Grenfell, B. T., & Baker, S. (2015). The impact of environmental and climatic variation on the spatiotemporal trends of hospitalized pediatric diarrhea in Ho Chi Minh City, Vietnam. *Health & Place*, 35, 147–154. https://doi.org/10.1016/j.healthplace.2015.08.001

Thornton, P. E., Running, S. W., & White, M. A. (1997). Generating surfaces of daily meteorological variables over large regions of complex terrain. *Journal of Hydrology*, 190(3–4), 214–251. https://doi.org/10.1016/S0022-1694(96)03128-9

Thu, H., Aye, K., & Thein, S. (1998). The effect of temperature and humidity on dengue virus propagation in Aedes aegypti mosquitos. *The Southeast Asian Journal of Tropical Medicine and Public Health*, 29(2), 280–284.

Thurston-Enriquez, J. A., Gilley, J. E., & Eghball, B. (2005). Microbial quality of runoff following land application of cattle manure and swine slurry. *Journal of Water and Health*, 3(2), 157–171. https://pubmed.ncbi.nlm.nih.gov/16075941/

Tien, J. H., & Earn, D. J. D. (2010). Multiple Transmission Pathways and Disease Dynamics in a Waterborne Pathogen Model. *Bulletin of Mathematical Biology*, 72(6), 1506–1533. https://doi.org/10.1007/s11538-010-9507-6

Tinker, S. C., Moe, C. L., Klein, M., Flanders, W. D., Uber, J., Amirtharajah, A., Singer, P., & Tolbert, P. E. (2010). Drinking water turbidity and emergency department visits for gastrointestinal illness in Atlanta, 1993-2004. *Journal of Exposure Science and Environmental Epidemiology*, 20(1), 19–28. https://doi.org/10.1038/jes.2008.68

Tinker, S. C., Moe, C. L., Klein, M., Flanders, W. D., Uber, J., Hall, A. B., Tolbert, P. E., Amirtharajah, A., Singer, P., & Tolbert, P. E. (2009). Drinking water residence time in distribution networks and emergency department visits for gastrointestinal illness in Metro Atlanta, Georgia. *Journal of Water and Health*, 7(2), 332–343. https://doi.org/10.2166/wh.2009.022

Tippett, R. (2016). The persistent "rurality" of North Carolina. In Carolina Demographics. Carolina Population Center, University of North Carolina. https://www.ncdemography.org/2016/03/21/the-persistent-rurality-of-north-carolina/

Tobías, A., & Saez, M. (2004). Time-series regression models to study the short-term effects of environmental factors on health. http://www3.udg.edu/fcee/economia/n11.pdf

Tonnang, H. E. Z., Kangalawe, R. Y. M., & Yanda, P. Z. (2010). Predicting and mapping malaria under climate change scenarios: the potential redistribution of malaria vectors in Africa. *Malaria Journal*, 9, 1–10. https://doi.org/10.1186/1475-2875-9-111

Tornevi, A., Axelsson, G., & Forsberg, B. (2013). Association between precipitation upstream of a drinking water utility and nurse advice calls relating to acute gastrointestinal illnesses. *PLoS ONE*, 8(7), e69918. https://doi.org/10.1371/journal.pone.0069918

Tornevi, A., Barregård, L., & Forsberg, B. (2015). Precipitation and primary health care visits for gastrointestinal illness in Gothenburg, Sweden. *PLOS ONE*, 10(5), e0128487. https://doi.org/10.1371/journal.pone.0128487

Tornevi, A., Bergstedt, O., & Forsberg, B. (2014). Precipitation effects on microbial pollution in a river: lag structures and seasonal effect modification. PLOS ONE, 9(5), e98546. https://doi.org/10.1371/journal.pone.0098546

Tornevi, A., Simonsson, M., Forsberg, B., Säve-Söderbergh, M., & Toljander, J. (2016). Efficacy of water treatment processes and endemic gastrointestinal illness - A multi-city study in Sweden. Water Research, 102, 263–270. https://doi.org/10.1016/j.watres.2016.06.018

Trærup, S. L. M., Ortiz, R. A., & Markandya, A. (2011). The costs of climate change: A study of cholera in Tanzania. International Journal of Environmental Research and Public Health, 8(12), 4386–4405. https://doi.org/10.3390/ijerph8124386

Trelease, A. W. (2006). Piedmont Urban Crescent. In W. S. Powell (Ed.), Encyclopedia of North Carolina. University of North Carolina Press. https://www.ncpedia.org/piedmont-urban-crescent

Troeger, C., Forouzanfar, M., Rao, P. C., Khalil, I., Brown, A., Reiner, R. C., Fullman, N., Thompson, R. L., Abajobir, A., Ahmed, M., Alemayohu, M. A., Alvis-Guzman, N., Amare, A. T., Antonio, C. A., Asayesh, H., Avokpaho, E., Awasthi, A., Bacha, U., Barac, A., … Mokdad, A. H. (2017). Estimates of global, regional, and national morbidity, mortality, and aetiologies of diarrhoeal diseases: a systematic analysis for the Global Burden of Disease Study 2015. The Lancet Infectious Diseases, 17(9), 909–948. https://doi.org/10.1016/S1473-3099(17)30276-1

Uejio, C. K., Yale, S. H., Malecki, K., Borchardt, M. A., Anderson, H. A., & Patz, J. A. (2014). Drinking water systems, hydrology, and childhood gastrointestinal illness in central and northern Wisconsin. *American Journal of Public Health*, 104(4), 639–646. https://doi.org/10.2105/AJPH.2013.301659

Upperman, C. R., Parker, J., Jiang, C., He, X., Murtugudde, R., & Sapkota, A. (2015). Frequency of extreme heat event as a surrogate exposure metric for examining the human health effects of climate change. *PLOS ONE*, 10(12), e0144202. https://doi.org/10.1371/journal.pone.0144202

U.S. Census Bureau. (2017). Vintage 2016 Population Estimates. https://www.census.gov/data/tables/2016/demo/popest/state-total.html

U.S. Census Bureau. (2021). Decennial Census of Population and Housing by Decades. 2020 Census of Population and Housing, Updated Every 10 Years. https://www.census.gov/programs-surveys/decennial-census/decade.2020.html

U.S. Census Bureau. (2022). Annual Estimates of the Resident Population for Selected Age Groups by Sex for North Carolina: April 1, 2020 to July 1, 2021 (SC-EST2021-AGESEX-37). U.S. Census Bureau, Population Division. https://www2.census.gov/programs-surveys/popest/tables/2020-2021/state/detail/sc-est2021-agesex-37.xlsx

US Environmental Protection Agency (EPA). (n.d.). Private Drinking Water Wells. Retrieved July 21, 2022, from https://www.epa.gov/privatewells

US Environmental Protection Agency (EPA). (2022). National Primary Drinking Water Regulations. https://www.epa.gov/ground-water-and-drinking-water/national-primary-drinking-water-regulations

U.S. GAO. (2008). Concentrated animal feeding operations: EPA needs more information and a clearly defined strategy to protect air and water quality from pollutants of concern (GAO-08-

944). In Environmental Concerns in Animal Agriculture and Feeding Operations (Issue September). https://www.gao.gov/new.items/d08944.pdf

USDA. (2007). The 2007 Census of Agriculture: Hogs and Pigs Fact Sheet. U.S. Department of Agriculture (USDA). https://www.agcensus.usda.gov/Publications/2007/Online_Highlights/Fact_Sheets/Production/hogsandpigs.pdf

USDA-ERS. (2022). State Fact Sheets: North Carolina. Economic Research Service, U.S. Department of Agriculture. https://data.ers.usda.gov/reports.aspx?StateFIPS=37&StateName=North Carolina&ID=17854

USDHHS, CDC, & CMS. (2009). ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification. U.S. Department of Health and Human Services (USDHHS), Centers for Disease Control and Prevention (CDC), Centers for Medicare and Medicaid Services (CMS).

USEPA-OW. (2013). Literature Review of Contaminants in Livestock and Poultry Manure and Implications for Water Quality (EPA 820-R-13-002; p. 943). U.S. Environmental Protection Agency (USEPA), Office of Water (OW). https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=P100H2NI.txt

Valipour, M. (2015). Long-term runoff study using SARIMA and ARIMA models in the United States. Meteorological Applications, 22(3), 592–598. https://doi.org/10.1002/met.1491

van de Schoot, R., Depaoli, S., King, R., Kramer, B., Märtens, K., Tadesse, M. G., Vannucci, M., Gelman, A., Veen, D., Willemsen, J., & Yau, C. (2021). Bayesian statistics and modelling. *Nature Reviews Methods Primers*, 1(1), 1. https://doi.org/10.1038/s43586-020-00001-2

Verbyla, D. L. (1987). Classification trees: a new discrimination tool. *Canadian Journal of Forest Research*, 17(9), 1150–1152. https://doi.org/10.1139/x87-177

Vitro, K. A., BenDor, T. K., Jordanova, T. v, & Miles, B. (2017). A geospatial analysis of land use and stormwater management on fecal coliform contamination in North Carolina streams. *Science of the Total Environment*, 603–604, 709–727. https://doi.org/10.1016/j.scitotenv.2017.02.093

von Hippel, P., Hunter, D., & Drown, M. (2017). Better Estimates from Binned Income Data: Interpolated CDFs and Mean-Matching. *Sociological Science*, 4, 641–655. https://doi.org/10.15195/v4.a26

Waddington, H., & Snilstveit, B. (2009). Effectiveness and sustainability of water, sanitation, and hygiene interventions in combating diarrhoea. *Journal of Development Effectiveness*, 1(3), 295–335. https://doi.org/10.1080/19439340903141175

Wade, T. J., Lin, C. J., Jagai, J. S., & Hilborn, E. D. (2014). Flooding and emergency room visits for gastrointestinal illness in Massachusetts: A case-crossover study. PLoS ONE, 9(10), e110474. https://doi.org/10.1371/journal.pone.0110474

Wakefield, J. (2007). Disease mapping and spatial regression with count data. Biostatistics, 8(2), 158–183. https://doi.org/10.1093/biostatistics/kxl008

Walker, K. (2016). tigris: An R Package to Access and Work with Geographic Data from the US Census Bureau. The R Journal, 8(2), 231. https://doi.org/10.32614/RJ-2016-043

Walker, K. (2023). Analyzing US Census Data: Methods, Maps, and Models in R (1st ed.). CRC Press.

Walker, K., & Herman, M. (2020). tidycensus: Load US Census Boundary and Attribute Data as "tidyverse" and "sf"-Ready Data Frames.

Wallender, E. K., Ailes, E. C., Yoder, J. S., Roberts, V. A., & Brunkard, J. M. (2014). Contributing Factors to Disease Outbreaks Associated with Untreated Groundwater. Groundwater, 52(6), 886–897. https://doi.org/10.1111/gwat.12121

Waller, A. E., Scholer, M., Ising, A. I., & Travers, D. A. (2011). Using emergency department data for biosurveillance: The North Carolina experience. In C. Castillo-Chavez, H. Chen, W. B. Lober, M. Thurmond, & D. Zeng (Eds.), Infectious Disease Informatics and Biosurveillance. Integrated Series in Information Systems, vol 27 (pp. 45–66). Springer US. https://doi.org/10.1007/978-1-4419-6892-0_3

Wang, H., Di, B., Zhang, T. J., Lu, Y., Chen, C., Wang, D., Li, T., Zhang, Z., & Yang, Z. (2019). Association of meteorological factors with infectious diarrhea incidence in Guangzhou, southern China: A time-series study (2006–2017). Science of the Total Environment, 672, 7–15. https://doi.org/10.1016/j.scitotenv.2019.03.330

Wang, H., Liu, Z., Xiang, J., Tong, M. X., Lao, J., Liu, Y., Zhang, J., Zhao, Z., Gao, Q., Jiang, B., & Bi, P. (2021). Effect of ambient temperatures on category C notifiable infectious diarrhea in China: An analysis of national surveillance data. Science of the Total Environment, 759(44), 143557. https://doi.org/10.1016/j.scitotenv.2020.143557

Wang, P., Goggins, W. B., & Chan, E. Y. Y. (2018b). A time-series study of the association of rainfall, relative humidity and ambient temperature with hospitalizations for rotavirus and norovirus infection among children in Hong Kong. Science of the Total Environment, 643, 414–422. https://doi.org/10.1016/j.scitotenv.2018.06.189

Wang, P., Goggins, W. B., & Chan, E. Y. Y. (2018a). Associations of Salmonella hospitalizations with ambient temperature, humidity and rainfall in Hong Kong. Environment International, 120, 223–230. https://doi.org/10.1016/j.envint.2018.08.014

Wang, P., Zhang, X., Hashizume, M., Goggins, W. B., & Luo, C. (2021). A systematic review on lagged associations in climate–health studies. International Journal of Epidemiology, 50(4), 1199–1212. https://doi.org/10.1093/ije/dyaa286

Wang, Y., Song, Q., Du, Y., Wang, J., Zhou, J., Du, Z., & Li, T. (2019a). A random forest model to predict heatstroke occurrence for heatwave in China. Science of the Total Environment, 650, 3048–3053. https://doi.org/10.1016/j.scitotenv.2018.09.369

Wang, Y., Song, Q., Du, Y., Wang, J., Zhou, J., Du, Z., & Li, T. (2019b). A random forest model to predict heatstroke occurrence for heatwave in China. Science of the Total Environment, 650, 3048–3053. https://doi.org/10.1016/j.scitotenv.2018.09.369

Wangdi, K., & Clements, A. C. (2017). Spatial and temporal patterns of diarrhoea in Bhutan 2003–2013. BMC Infectious Diseases, 17(1), 507. https://doi.org/10.1186/s12879-017-2611-6

Wangdi, K., Singhasivanon, P., Silawan, T., Lawpoolsri, S., White, N. J., & Kaewkungwal, J. (2010a). Development of temporal modelling for forecasting and prediction of malaria infections using time-series and ARIMAX analyses: A case study in endemic districts of Bhutan. Malaria Journal, 9(1), 251. https://doi.org/10.1186/1475-2875-9-251

Wangdi, K., Singhasivanon, P., Silawan, T., Lawpoolsri, S., White, N. J., & Kaewkungwal, J. (2010b). Development of temporal modelling for forecasting and prediction of malaria infections using time-series and ARIMAX analyses: A case study in endemic districts of Bhutan. Malaria Journal, 9(1), 251. https://doi.org/10.1186/1475-2875-9-251

Watson, J., Gayer, M., & Connolly, M. (2006). Epidemic risk after disasters. Emerging Infectious Diseases, 12(7), 1468–1469. https://doi.org/10.3201/eid1209.060500

Wendland, W. M. (1987). Hydroclimatology. In Climatology (pp. 497–502). Kluwer Academic Publishers. https://doi.org/10.1007/0-387-30749-4_90

White, A. N. J., Kinlin, L. M., Johnson, C., Spain, C. V., Ng, V., & Fisman, D. N. (2009). Environmental determinants of campylobacteriosis risk in Philadelphia from 1994 to 2007. EcoHealth, 6(2), 200–208. https://doi.org/10.1007/s10393-009-0246-9

Wikle, C. K., Berliner, L. M., & Cressie, N. (1998). Hierarchical Bayesian space-time models. Environmental and Ecological Statistics, 5(2), 117–154. https://doi.org/10.1023/A:1009662704779

Wing, S., Cole, D., & Grant, G. (2000). Environmental injustice in North Carolina's hog industry. Environmental Health Perspectives, 108(3), 225–231. https://doi.org/10.1289/ehp.00108225

Wing, S., Freedman, S., & Band, L. (2002). The potential impact of flooding on confined animal feeding operations in eastern North Carolina. Environmental Health Perspectives, 110(4), 387–391. https://doi.org/10.1289/ehp.02110387

Wing, S., & Johnston, J. (2014). Industrial Hog Operations in North Carolina Disproportionately Impact African-Americans, Hispanics and American Indians. Title IV Complaint. http://www.ncpolicywatch.com/wp-content/uploads/2014/09/UNC-Report.pdf

Wing, S., & Wolf, S. (2000). Intensive livestock operations, health, and quality of life among eastern North Carolina residents. Environmental Health Perspectives, 108(3), 233–238. https://doi.org/10.1289/ehp.00108233

WMO. (2009). Guidelines on analysis of extremes in a changing climate in support of informed decisions for adaptation. In A. M. G. Klein Tank, F. W. Zwiers, & X. Zhang (Eds.), WCDMP No. 72, WMO-TD No. 1500.

Wolf, J., Prüss-Ustün, A., Cumming, O., Bartram, J., Bonjour, S., Cairncross, S., Clasen, T., Colford, J. M., Curtis, V., de France, J., Fewtrell, L., Freeman, M. C., Gordon, B., Hunter, P. R., Jeandron, A., Johnston, R. B., Mäusezahl, D., Mathers, C., Neira, M., & Higgins, J. P. T. (2014). Assessing the impact of drinking water and sanitation on diarrhoeal disease in low- and middle-income settings: systematic review and meta-regression. *Tropical Medicine & International Health*, 19(8), 928–942. https://doi.org/10.1111/tmi.12331

Wong, D. (2009). The Modifiable Areal Unit Problem (MAUP). In A. S. Fotheringham & P. A. Rogerson (Eds.), *The SAGE Handbook of Spatial Analysis* (Issue 7, pp. 104–123). SAGE Publications, Ltd. https://doi.org/10.4135/9780857020130.n7

Wong, D. W. S. (2004). The Modifiable Areal Unit Problem (MAUP). In D. G. Janelle, B. Warf, & K. Hansen (Eds.), *WorldMinds: Geographical Perspectives on 100 Problems* (pp. 571–575). Springer Netherlands. https://doi.org/10.1007/978-1-4020-2352-1_93

Wood, S. N. (2001). mgcv: GAMs and Generalized Ridge Regression for R. *R News*, 1(2), 20–25.

Wood, S. N. (2004). Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association*, 99(467), 673–686. https://doi.org/10.1198/016214504000000980

Wood, S. N. (2006). *Generalized Additive Models: An Introduction with R (1st ed.)*. Chapman & Hall/CRC.

Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R (2nd ed.)*. Chapman and Hall/CRC. https://doi.org/10.1201/9781315370279

Wood, S. N. (2020). Package "mgcv." In CRAN (pp. 1–476).

World Health Organization (WHO). (2003). Climate change and human health: Risks and responses (A. J. McMichael, D. H. Campbell-Lendrum, C. F. Corvalan, K. L. Ebi, A. K. Githeko, J. D. Scheraga, & A. Woodward, Eds.). World Health Organization (WHO).

World Health Organization (WHO). (2017). Diarrheal Diseases. http://www.who.int/news-room/fact-sheets/detail/diarrhoeal-disease

Wu, J., Yunus, M., Ali, M., Escamilla, V., & Emch, M. (2018). Influences of heatwave, rainfall, and tree cover on cholera in Bangladesh. *Environment International*, 120(3), 304–311. https://doi.org/10.1016/j.envint.2018.08.012

Wu, J., Yunus, M., Streatfield, P. K., & Emch, M. (2014). Association of climate variability and childhood diarrhoeal disease in rural Bangladesh, 2000-2006. *Epidemiology and Infection*, 142(9), 1859–1868. https://doi.org/10.1017/S095026881300277X

Wu, X., Lu, Y., Zhou, S., Chen, L., & Xu, B. (2016). Impact of climate change on human infectious diseases: Empirical evidence and human adaptation. *Environment International*, 86, 14–23. https://doi.org/10.1016/j.envint.2015.09.007

Xie, P., Joyce, R., Wu, S., Yoo, S.-H., Yarosh, Y., Sun, F., Lin, R., & NOAA CDR Program. (2019). NOAA Climate Data Record (CDR) of CPC Morphing Technique (CMORPH) High Resolution Global Precipitation Estimates, Version 1. https://doi.org/10.25921/w9va-q159

Xu, C., Xiao, G., Wang, J., Zhang, X., & Liang, J. (2018). Spatiotemporal risk of bacillary dysentery and sensitivity to meteorological factors in Hunan Province, China. *International Journal of Environmental Research and Public Health*, 15(1), 47. https://doi.org/10.3390/ijerph15010047

Xu, Z., Hu, W., Zhang, Y., Wang, X., Zhou, M., Su, H., Huang, C., Tong, S., & Guo, Q. (2015). Exploration of diarrhoea seasonality and its drivers in China. *Scientific Reports*, 5(1), 8241. https://doi.org/10.1038/srep08241

Xu, Z., Huang, C., Turner, L. R., Su, H., Qiao, Z., & Tong, S. (2013). Is diurnal temperature range a risk factor for childhood diarrhea? *PLoS ONE*, 8(5), e64713. https://doi.org/10.1371/journal.pone.0064713.t001

Xu, Z., Liu, Y., Ma, Z., Toloo, G., Hu, W., & Tong, S. (2014). Assessment of the temperature effect on childhood diarrhea using satellite imagery. *Scientific Reports*, 4, 1–8. https://doi.org/10.1038/srep05389

Yan, L., Wang, H., Zhang, X., Li, M.-Y., & He, J. (2017). Impact of meteorological factors on the incidence of bacillary dysentery in Beijing, China: A time series analysis (1970-2012). *PLOS ONE*, 12(8), e0182937. https://doi.org/10.1371/journal.pone.0182937

Yang, K., LeJeune, J., Alsdorf, D., Lu, B., Shum, C. K., & Liang, S. (2012). Global Distribution of Outbreaks of Water-Associated Infectious Diseases. *PLoS Neglected Tropical Diseases*, 6(2), e1483.

Yun, J., Greiner, M., Höller, C., Messelhäusser, U., Rampp, A., & Klein, G. (2016). Association between the ambient temperature and the occurrence of human Salmonella and Campylobacter infections. *Scientific Reports*, 6(1), 28442. https://doi.org/10.1038/srep28442

Zambrano, L. D., Levy, K., Menezes, N. P., & Freeman, M. C. (2014). Human diarrhea infections associated with domestic animal husbandry: A systematic review and meta-analysis.

Transactions of The Royal Society of Tropical Medicine and Hygiene, 108(6), 313–325. https://doi.org/10.1093/trstmh/tru056

Zanobetti, A., Wand, M. P., Schwartz, J., & Ryan, L. M. (2000). Generalized additive distributed lag models: quantifying mortality displacement. *Biostatistics*, 1(3), 279–292. https://doi.org/10.1093/biostatistics/1.3.279

Zeger, S. L., & Liang, K.-Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, 42(1), 121–130. https://doi.org/10.2307/2531248

Zhang, X., Alexander, L., Hegerl, G. C., Jones, P., Tank, A. K., Peterson, T. C., Trewin, B., & Zwiers, F. W. (2011). Indices for monitoring changes in extremes based on daily temperature and precipitation data. *Wiley Interdisciplinary Reviews: Climate Change*, 2(6), 851–870. https://doi.org/10.1002/wcc.147

Zhang, X., Gu, X., Wang, L., Zhou, Y., Huang, Z., Xu, C., & Cheng, C. (2021). Spatiotemporal variations in the incidence of bacillary dysentery and long-term effects associated with meteorological and socioeconomic factors in China from 2013 to 2017. *Science of the Total Environment*, 755, 142626. https://doi.org/10.1016/j.scitotenv.2020.142626

Zhang, Y., Bi, P., & Hiller, J. (2008a). Climate variations and salmonellosis transmission in Adelaide, South Australia: a comparison between regression models. *International Journal of Biometeorology*, 52(3), 179–187. https://doi.org/10.1007/s00484-007-0109-4

Zhang, Y., Bi, P., & Hiller, J. E. (2008b). Weather and the transmission of bacillary dysentery in Jinan, Northern China: A time-series analysis. *Public Health Reports*, 123(1), 61–66. https://doi.org/10.1177/003335490812300109

Zhang, Y., Bi, P., & Hiller, J. E. (2010). Climate variations and Salmonella infection in Australian subtropical and tropical regions. *Science of the Total Environment*, 408(3), 524–530. https://doi.org/10.1016/j.scitotenv.2009.10.068

Zhang, Y., Bi, P., Hiller, J. E., Sun, Y., & Ryan, P. (2007). Climate variations and bacillary dysentery in northern and southern cities of China. *Journal of Infection*, 55(2), 194–200. https://doi.org/10.1016/j.jinf.2006.12.002

Zhou, X., Zhou, Y., Chen, R., Ma, W., Deng, H., & Kan, H. (2013). High temperature as a risk factor for infectious diarrhea in Shanghai, China. *Journal of Epidemiology*, 23(6), 418–423. https://doi.org/10.2188/jea.JE20130012

Zmirou, D., Rey, S., Courtois, X., Ferley, J.-P., Blatier, J.-F., Chevallier, P., Boudot, J., Potelon, J.-L., & Mounir, R. (1995). Residual microbiological risk after simple chlorine treatment of drinking ground water in small community systems. *The European Journal of Public Health*, 5(2), 75–81. https://doi.org/10.1093/eurpub/5.2.75