“MULTI-SCALE DRIVERS OF INFECTIOUS DISEASE TRANSMISSION AND MORTALITY, AND IMPLICATIONS FOR CONTROL AND PREVENTION”

by

RACHEL MERCALDO

(Under the Direction of Andreas Handel)

ABSTRACT

From molecular-level determinants such as host and pathogen genetics to ecosystem-scale factors such as spatial distributions and climate patterns, the determinants of infectious diseases are highly variable. Due to this biocomplexity, researchers have called for a more comprehensive view of infectious diseases, their determinants and affected populations. An “across-scales” perspective of infectious disease epidemiology acknowledges the role of each biological level of organization in the complexity of overall disease distribution, while maintaining a focus on human population health. The aims of this dissertation represent work toward understanding the transmission, severity, or prevention of three different infectious pathogens using such an approach: SARS-CoV-2, *Mycobacterium tuberculosis*, and Zika virus. While each disease can itself be studied across multiple scales, these studies focus on specific biological levels of organization. COVID-19 mortality is modeled using spatiotemporal methods, illustrating the effect of community and ecosystem level dynamics. Population and community social and medical contacts are examined for patterns associated with delay to diagnosis of pulmonary tuberculosis disease. Finally, individual, population, and ecosystem dynamics are included in a simulation study of a Zika virus vaccine clinical trial. Together, these projects

highlight some of the biological scales contributing to the epidemiology of these emerging or reemerging infectious diseases, and illustrate the broader necessity for an across-scales framework in infectious disease research.

INDEX WORDS: Mycobacterium tuberculosis, SARS-CoV-2, COVID-19, Zika virus, Congenital Zika Syndrome, diagnostic delay, clinical trials, vaccination, environmental models, spatiotemporal models

MULTI-SCALE DRIVERS OF INFECTIOUS DISEASE TRANSMISSION AND MORTALITY, AND IMPLICATIONS FOR CONTROL AND PREVENTION

by

RACHEL MERCALDO

BA, Pennsylvania State University, 2014

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2021

MULTI-SCALE DRIVERS OF INFECTIOUS DISEASE TRANSMISSION AND MORTALITY, AND IMPLICATIONS FOR CONTROL AND PREVENTION

by

RACHEL MERCALDO

Major Professor:	Andreas Handel
Committee:	José Cordero
	Juliet Sekandi
	Daniel Mead

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
May 2021

ACKNOWLEDGEMENTS

Thank you to Dr. Andreas Handel for your mentorship during the dissertation process. Thank you to Dr. Vanessa Ezenwa, Dr. Andrew Park, Dr. Sonia Altizer, Dr. Richard Hall, and other Odum School of Ecology faculty for making the Interdisciplinary Disease Ecology Across Scales (IDEAS) program possible, and for your support during my graduate education. Thank you to Dr. Daniel Mead, Dr. José Cordero, and Dr. Juliet Sekandi for serving on my committee and for your patience and assistance. Special thanks to Dr. Steven Bellan, for your guidance in my first years in the program. Finally, thank you to my family and friends, in Athens and beyond, without whom this dissertation would not be possible.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS iv

LIST OF TABLES vii

LIST OF FIGURES viii

CHAPTER

1 INTRODUCTION AND LITERATURE REVIEW 1

2 EFFECT OF ENVIRONMENTAL AND COMMUNITY VARIABLES ON CLUSTERS OF COVID-19 MORTALITY IN US COUNTIES 5

3 COMMUNITY DRIVERS OF TUBERCULOSIS DIAGNOSTIC DELAY IN KAMPALA, UGANDA 19

4 EVALUATION OF ALTERNATIVE ENDPOINTS FOR ZIKV VACCINE EFFICACY TRIALS 38

5 CONCLUSION 56

REFERENCES 59

LIST OF TABLES

Page

Table 2.1: Characteristics of counties within and outside of reported clusters 17

Table 2.2: Details of clustering timing, observed deaths, and expected deaths.....18

Table 2.3: Negative binomial model (log link) results 18

Table 3.1: Baseline characteristics of participants in the Diagnostic II study.....33

Table 3.2: Results of bivariate linear regression for significant ($p < 0.05$) predictors 34

Table 3.3: Results of linear regression (with LASSO regularization) models of community.....35

Table 3.4: Results of bivariate linear regression for significant ($p < 0.05$) predictors 36

Table 3.5: Results of linear regression (with LASSO regularization) models of contribution.....37

Table 4.1: Simulation parameters 54

Table 4.2: Minimum required sample size and study duration for trial scenarios with 80% 55

LIST OF FIGURES

Page

Figure 2.1: Example of temperature (left) and relative humidity (right) imputation from.....10

Figure 2.2: Spatial distribution of selected covariates.....12

Figure 2.3: Google Earth imaging of SaTScan space-time permutation model results.....13

Figure 2.4: Coefficients of negative binomial regression.....14

Figure 3.1: Conceptual framework, with study definitions and an example of a realized 22

Figure 3.2: Community contact delay.....27

Figure 4.1: Scaled person-week infection rates for the six countries included in the study.....43

Figure 4.2: Statistical power to detect vaccine efficacy by number of recruited participants.....48

Figure 4.3: Number of infections occurring in the first four weeks of simulated trials 49

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Statement of the Problem

Epidemiology is defined as the study of the distribution and determinants of diseases—or other health-related states or outcomes—in populations, and the application of such study to prevention and control efforts [1]. Key to this definition is the specification of a target population, that group of individuals about whom conclusions are reported [2]. An epidemiologist is concerned with the frequency and patterns of disease in a group of individuals, rather than the health of a single person. This focus immediately provides an infectious disease epidemiologist with the *distribution* portion of the basic definition of the field [2]. The *determinants* of a disease can then be determined by comparing disease rates between groups experiencing varying levels of exposure [2]. Males are more likely than females to activate to symptomatic tuberculosis disease [3]. Elderly individuals are more likely to suffer severe outcomes of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection [4]. Pregnant women, when infected with Zika virus (ZIKV), are at risk of pregnancy loss, while the Congenital Zika Syndrome risk to the fetus depends on timing of infection in the course of pregnancy [5]. Many questions about infectious disease determinants can be answered with frequency or distributions data. Despite this, researchers have called for a more comprehensive view of infectious diseases, their causes and affected populations, citing the complexity of transmission and virulence that depends on factors at multiple scales of organization [6]. From

molecular–level determinants such as host and pathogen genetics to ecosystem-scale factors such as spatial distributions and climate patterns, the determinants of disease are highly variable [6].

An “across-scales” perspective of infectious disease epidemiology acknowledges the role of each biological level of organization in the complexity of overall disease distribution, while maintaining a focus on human population health. While the results of such studies still apply to target populations, contributing biological scales range from molecular through cellular, organismal, population, community, and ecosystem levels.

Though disparate diseases, the three pathogens used as examples, above, are similar in that they follow a pattern of emergence or reemergence with complex underlying mechanisms [6]. SARS-CoV-2 and ZIKV are both recently emerged viruses, while tuberculosis remains a leading cause of death despite monumental control and prevention efforts over decades. Wilcox and Colwell [7] state that understanding factors responsible for the emergence of new diseases, or a reemergence, is one of the most imperative issues facing society. These authors argue for a more holistic perspective of the crisis, incorporating social, physical, chemical, and biological dimensions [7].

The term “biocomplexity” was coined in this context by Colwell [8], who applied a comprehensive framework in the example of climate and cholera in 1996 [9]. The author cited the large spatial and temporal scale of multiple cholera pandemics, findings from a screening of preserved plankton samples, and patterns of El Niño events in her conclusions, illustrating the necessity of an interdisciplinary “cross-cut” of oceanography, microbiology, marine biology, ecology, medicine, epidemiology, and even satellite technology fields [9].

The spread of antibiotic resistance is an excellent example of more recent, interdisciplinary efforts to better understand a complex infectious disease problem from more

minute scales to those at a community-level. Chronic infections associated with resistant nosocomial pathogens are of particular concern, as they are associated with greater transmission when patients visit multiple healthcare locations or remain hospitalized for longer periods of time [10]. These bacteria have been found in environments surrounding hospitals, surviving in drains, pools, and rivers, where they present a threat to both domesticated animals and wildlife [11]. Veterinary practitioners use many of the same antibiotics used in human medicine, and even though there is a call that "last-resort" antibiotics be reserved for human infections, bacteria resistant to these antibiotics have been isolated from both domesticated animals and wildlife [12,13]. Wildlife contributions to the spread of resistance are expected to increase in this time of rapid urbanization [14].

While such in-depth evaluations of the importance of a multi-scale approach are not available for most infectious diseases, the previous examples highlight the value of cooperation between experts, and contributions from multiple fields to solve the problem of infectious disease emergence or spread. The aims of this dissertation represent work toward understanding the transmission or control efforts of three different infectious pathogens: severe acute respiratory syndrome coronavirus 2, *Mycobacterium tuberculosis*, and Zika virus. In each, analyses or simulations of data at different biological scales is highlighted, and the role of the scale of the data considered in the dissertation conclusion.

Specific Aims of the Dissertation

The primary goal of the dissertation is to better describe the transmission or control efforts of three pathogens, and to highlight the biological scales contributing to their spread or prevention. The aims are presented from the broadest to narrowest scale, beginning with the effect of climate and community factors on SARS-CoV-2 transmission and COVID-19 mortality, and concluding

with the effect of individual pregnancy probability and timing of Zika virus infection on the feasibility of a vaccine clinical trial against Congenital Zika Syndrome. The titles of the specific aims are:

1. "Effect of environmental and community variables on clusters of COVID-19 mortality in US counties: a spatiotemporal analysis."
2. "Community drivers of tuberculosis diagnostic delay in Kampala, Uganda: A retrospective cohort study."
3. "Evaluation of alternative endpoints for ZIKV vaccine efficacy trials."

Dissertation Outline

Chapter 1 of this dissertation as served as a brief introduction to the importance of across-scales infectious disease research. Chapter 2 through Chapter 4 describe each aim in detail, providing background, methodology, details of analysis, and results. Chapter 5 concludes the dissertation with a discussion of the across-scales components of each aim.

CHAPTER 2

EFFECT OF ENVIRONMENTAL AND COMMUNITY VARIABLES ON CLUSTERS OF COVID-19 MORTALITY IN US COUNTIES: A SPATIOTEMPORAL ANALYSIS[1]

[1] Mercaldo, R.A., Richards, R., Schmidt, J.P., Drake, J., and Handel, A. To be submitted to *Spatial and Spatio-temporal Epidemiology*.

Abstract

The COVID-19 pandemic has claimed over 2.5 million lives since late 2019. Its rapid geographic spread has spurred much research on transmission and risk factors, with several spatial or spatiotemporal studies identifying environmental and demographic factors as predictors of reported COVID-19 cases or deaths. In this study, we build on other work with a spatiotemporal model that explicitly includes mortality's dependence on previous transmission patterns. Using this model, we examine patterns of mortality to identify clusters of counties with higher or lower mortality than expected, given local transmission patterns, and explore factors associated with mortality and clustering. We discovered eight significant clusters, in space and time, and determine that any mortality clustering is explainable by known risk factors, including proportion of the population 65 years or older, the proportion of the population identifying as Black or Hispanic, the availability of ICU beds, and environmental factors.

Introduction

The novel coronavirus, Severe Acute Respiratory Syndrome Coronavirus 2, or SARS-CoV-2, has caused over 116 million confirmed cases and 2.5 million deaths from COVID-19 [15]. Unlike previous epidemic coronaviruses—severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East Respiratory Syndrome coronavirus (MERS-CoV)—SARS-CoV-2 is transmitted far more efficiently outside of healthcare settings, through contacts in the community [16], and quickly became a global pandemic after its identification in China in late 2019. Its rapid geographic spread spurred numerous studies focused on transmission, with a subset of these projects focused on spatially-explicit drivers.

An early review of geospatial studies, published in May 2020, had already identified 63 scientific articles [17], a large portion of which were dedicated to predictors with "environmental" or "health and social geography" themes. Few studies that accounted for both temporal and spatial variation were conducted in the United States. One GIS-based spatial modeling study found clusters of regions with significantly increased risk of transmission and mortality in nursing homes across the United States [18]. Higher proportions of Black individuals and non-English speakers were significant predictors of increased cases in spatial lag models [19]. The same study found that higher proportions of Black individuals and persons living with disability significantly predicted increased deaths [19]. Another study sought to produce a rapid surveillance method using space-time scan statistics to identify clustering of cases at the county level [20]. And while it is difficult to measure compliance with ordered control measures, studies have attempted to better understand local behavior and its relation to transmission. One analysis reported on social behavior related to political leaning and likelihood of social distancing [21].

Meteorological predictors, particularly temperature and humidity, have received a great deal of attention, having been implicated in reducing the viability of various other viruses in the environment, potentially influencing inoculum dose and transmission probability. A modeling study illustrating droplet evaporation and viral suspension of coronaviruses in air following normal breathing, coughing, or sneezing reported that higher humidity reduces evaporation rates sufficiently to increase falling time, potentially reducing airborne transmission [22], though this could consequently increase surface contamination. Evidence from the Beijing SARS-CoV outbreak linked such laboratory evidence to transmission: temporally-lagged average temperatures and humidity were significantly associated with incident SARS cases, with both

higher temperature and humidity negatively correlated with incidence [23]. The seasonal nature of previous coronavirus outbreaks, along with the small number of laboratory or epidemiological studies of these pathogens, led to a cautiously optimistic view that warm, humid summer months may dampen transmission of COVID-19. However, a simulation study by Baker, et al., highlighted that the effect of meteorological variables on transmission may be overshadowed by known epidemic drivers, such as high population susceptibility during the current pandemic stage of SARS-CoV-2 [24]. Indeed, previous studies reporting an effect of humidity or temperature on transmission have focused primarily on endemic pathogens, particularly seasonal influenza [25–28]. Studies on illness severity due to climatological drivers are rare and mainly involve animal models, but relative humidity in particular has been found associated with recovery from influenza in murine models [29]. Evidence in humans is lacking, though provision of humid air to hospitalized patients has been recommended [30], and the hydrating effect of masking has been proposed as a potential explanation for masks reducing disease severity [31]. While the effect of meteorological or other large-scale variables may be difficult to discern, the quick response and ongoing contributions from the scientific community have generated a tremendous amount of information about SARS-CoV-2, with numerous sources of data made publicly available. These resources have revealed a trend in the spatiotemporal literature of analyzing factors related to either cases (or other measures of transmission) or mortality. To our knowledge, no spatial study has explicitly included the dependent nature of mortality on transmission, or sought to identify local factors associated with mortality rates greater than expected given local transmission.

In our current analysis, we employ a county-level space-time autoregressive (STAR) framework to quantify the effect of environmental and community variables on the number of

new deaths reported each week. We examine patterns of mortality to identify clusters of counties with higher than expected mortality, given cases reported as a measure of local transmission patterns. We then explore factors associated with these clusters.

Methods

Data

COVID-19 Cases and Mortality.

Publicly available data on COVID-19 cases and mortality was obtained from the New York Times [32] Github repository. Data was reduced to include only the first thirty weeks of the US epidemic, or until the decrease in cases that began in mid-August, 2020 and generally continued until the holiday season. All counties in the contiguous US were included, while data from Alaska, Hawaii, the US Virgin Islands, Guam, and other noncontiguous locations were removed from the dataset.

If data showed a negative case number or negative number of deaths in any given week, we attempted to correct for errors in reporting by subtracting that number of deaths from the most recent week with non-zero counts, and assigning a 0 for the originally negative count. Any remaining negatives were set to 0.

Few cases were reported in the first five weeks of the epidemic data, with only three counties reporting cases in or before week five. We started analysis at week six to avoid excessive counts of zero deaths in the data set, though previous weeks' case counts were included at a lag to include transmission patterns in models.

Climatological Data.

US weather station data was downloaded from the Global Summary of the Day database of the National Oceanic and Atmospheric Association (NOAA) via the R package GSODR

[33,34]. Daily mean temperature (in degrees Celsius) and relative humidity were chosen as predictors and summarized into weekly averages for each weather station. There were a total of 1050 weather stations included in the dataset. For each week, the station data was used to impute temperature and relative humidity averages for a 300-pixel grid overlaying the US map, through kriging spatial imputation methods using the kriging R package [35]. County-level values were determined by averaging the imputed data for all grid points that fell within county borders. Figure 2.1 shows example kriging maps.

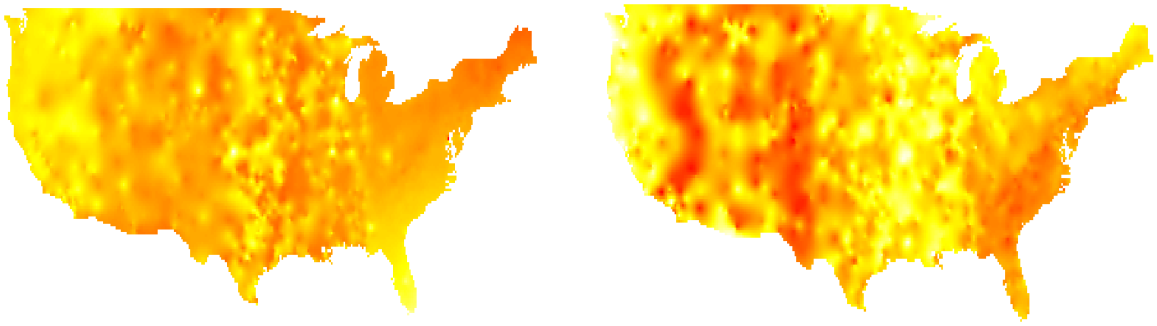

Figure 2.1 Example of temperature (left) and relative humidity (right) imputation from 1050 weather station data points.

Other Covariates.

County-level measures of population size and the number of people identifying as Black or Hispanic were recorded from the US Census Bureau [36]. Food insecurity data was used as a measure of socio-economic status, and was obtained from Feeding America [37]. Data on the number of ICU beds, aggregated by county, as well as population density was downloaded from ArcGIS [38]. Most covariates were available only for previous years: while ArcGIS updates the data on ICU beds per county, census data used is from 2010, other ArcGIS data from 2012, and Feeding America food insecurity data from 2018.

STAR Framework

The Space-Time AutoRegressive framework employed in this study associates the outcome of interest at one time point with that at a previous time point. For example, reported COVID-19 deaths in one week are expected to be related to death counts in the previous week, and this information is added to models to control for the temporal autocorrelation. Similarly, the outcome in one county can be expected to be related to the outcomes in nearby counties. To account for this spatial autocorrelation, the previous week's deaths in neighboring counties are averaged and added to our dataset as a predictor. We do not assume an instantaneous effect between counties, in that we do not include *current* deaths in the neighboring counties as a covariate.

We included cases reported three weeks prior to death reports, to account for the lag between case reporting and deaths. This assumes that deaths reported in one week result from cases reported three weeks previously.

We included environmental covariates at a one-week lag. County population size, population density, proportion of the population 65 years or older, and proportion of the population identifying as Black (either Black alone or Black with one other race) were assumed constant over the study period.

Analysis

A spatial scan statistic was applied to the time series of mortality, associated with county centroids, to identify clustering of mortality when accounting for the lagged time series of cases reported. A lag of three weeks was used. The spatial scan analysis was conducted in SaTScan [39], scanning for clusters of high rates using a retrospective space-time permutation model. The population demographic features of any counties included in clusters at any time during the

epidemic were compared to counties never included in clusters. We fit a zero-inflated binomial regression model (with logit link) with these constant county-level features, as well as temporally-variable environmental features, as covariates. Significant predictors in this model were then included as covariates in a final space-time permutation model, to identify any high-mortality clusters remaining after adjusting for covariates.

Results

Five significant clusters of high mortality, given cases reported three weeks prior, were identified using the space-time permutation model in SaTScan. In the same model, three significant clusters of lower than expected mortality were identified. Demographic features for counties included or not included in significant clusters are reported in Table 2.1. County feature maps are shown in Figure 2.2.

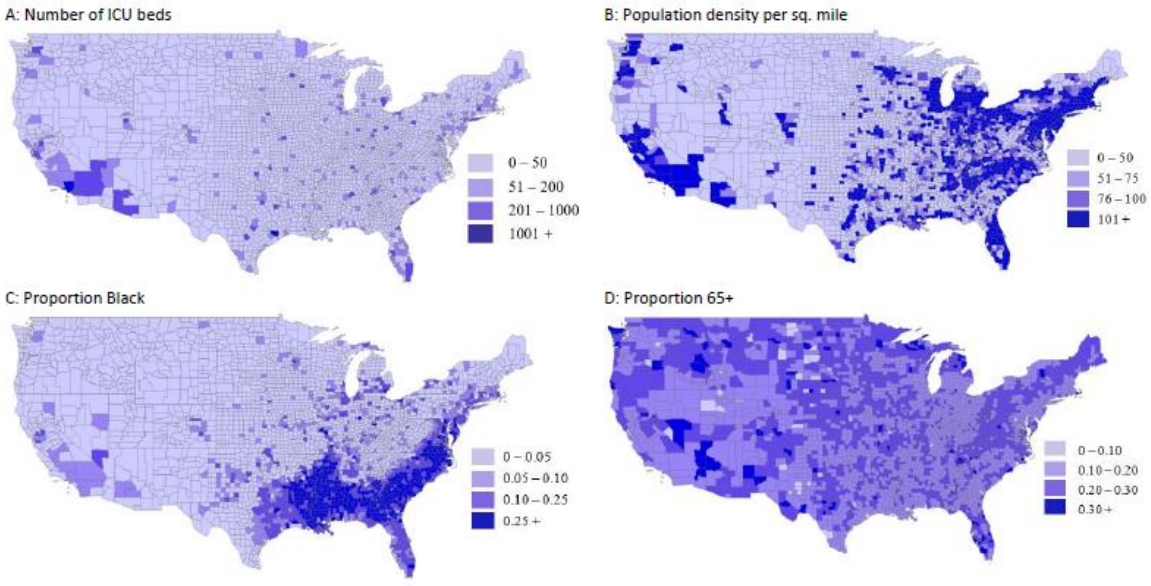

Figure 2.2. Spatial distribution of selected covariates.

Clusters were identified in both space and time. A visual of spatial clustering is included in Figure 2.3. All clusters were identified early in the epidemic, in March and April. Details of the time frame for each cluster, as well as the observed and expected numbers of deaths, are reported in Table 2.2.

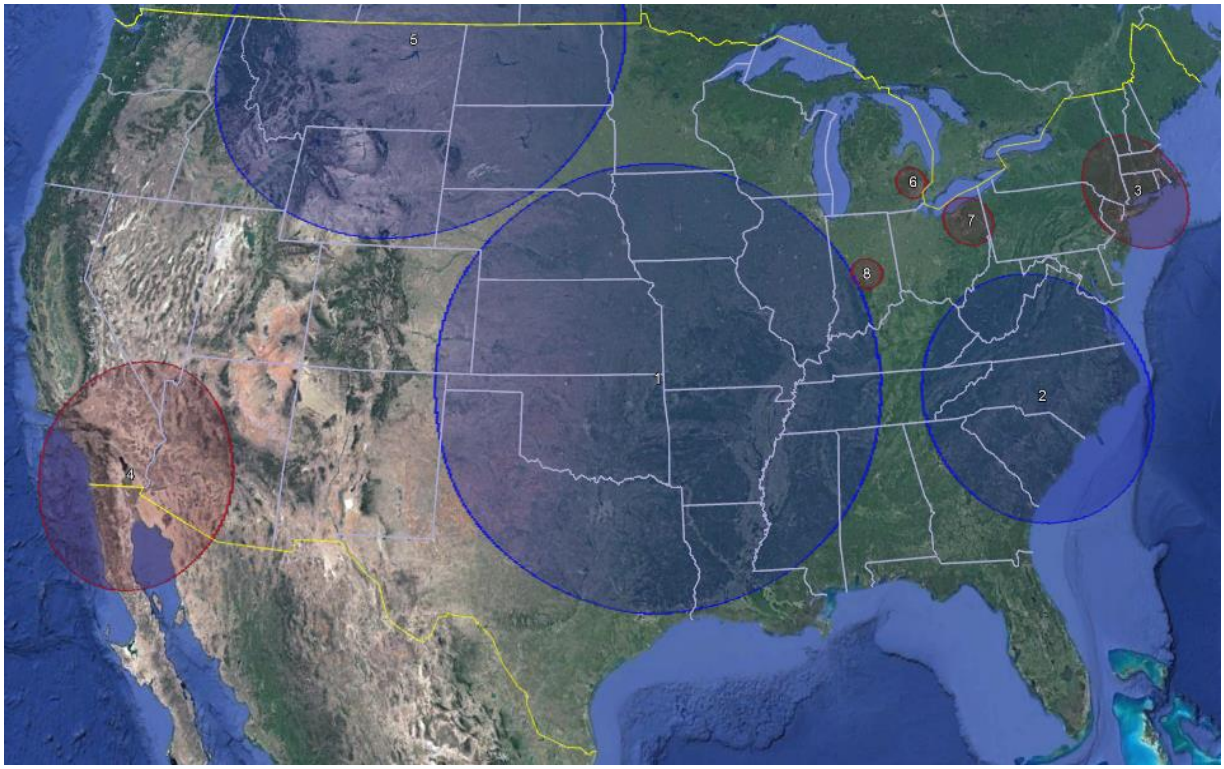

Figure 2.3: Google Earth © imaging of SaTScan space-time permutation model results, identifying significant (p-value < 0.05) clusters of higher (red) or lower (blue) weekly reported mortality than expected, after controlling for cases reported three weeks prior in base and neighboring counties as covariates. The radius of the circles indicates the size of the cluster (in km). *Attribution: US Dept of State Geographer © 2021 Google*

Negative binomial model and final space-time permutation model

We fit a negative binomial model with all predictors. Only the number of ICU beds in the county of interest's neighboring counties was not found significant (Table 2.3, Figure 2.4). We included all significant predictors in a final SaTScan space-time permutation model, to identify

any remaining clustering. The final space-time model identified no remaining clusters after controlling for these covariates.

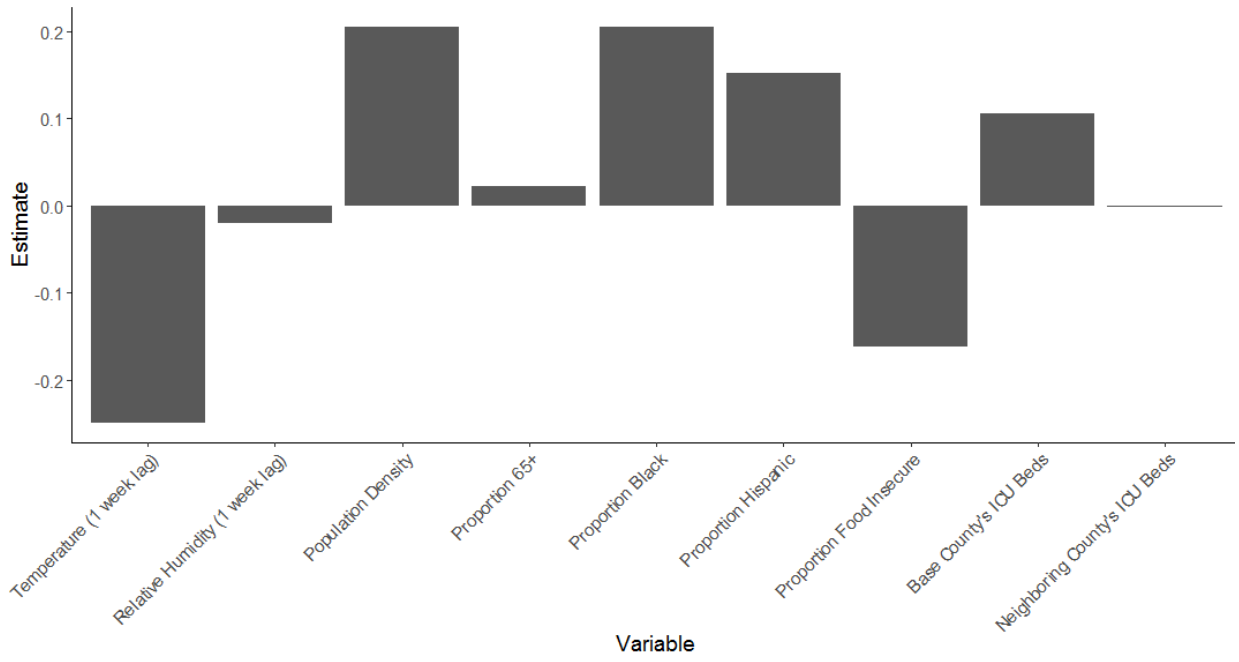

Figure 2.4: Coefficients of negative binomial regression.

Discussion

In this study, we identified significant clusters of higher and lower than expected mortality from COVID-19 in US counties, employing a space-time autoregressive (STAR) framework to control for transmission patterns. After adjusting the model with demographic and community-level covariates found significant in a separate negative binomial model, clustering was no longer evident, indicating known drivers of COVID-19 mortality are sufficient in predicting deaths at the county level.

The STAR framework controlled for both spatial and temporal auto-correlation in a zero-inflated model. Biologically-relevant temporal and spatial lags for other variables were also

calculated for the dataset: environmental variables associated with illness in animal models, as well as neighboring county ICU beds in addition to those in the county of interest. Population density, the proportion of the population 65 years or older, proportion of the population identifying as Black or Hispanic alone, and the number of ICU beds in a county are associated with increased mortality, while temperature, humidity, and food insecurity are associated with lower mortality.

Notably, food insecurity, like most other covariates, was available only for previous years. While many county-level features do not change significantly from year to year, food insecurity and other socio-economic features are likely impacted by the COVID-19 pandemic and subsequent workplace closures and stay-at-home orders. A report by Feeding America estimates the effect of COVID-19 on unemployment, poverty, and food insecurity, and makes general projections [37].

As the quality of any study is affected by the accuracy of data, errors in reported deaths and cases of COVID-19 represent a further limitation. Space-time permutation models have been found to perform well, however, even in cases with poor spatial and temporal accuracy [40].

The data used in this study was aggregated to a weekly, county level. The space-time permutation model used controls for both spatial and temporal trends, comparing the number of deaths reported in a cluster to what would have been expected if counties were independent of each other in space and time. Clusters thus represent specific areas, in a specific time period, when the number of deaths is greater than or lower than expected when areas outside of the cluster experience an expected numbers of deaths [39]. One limitation of this method is that it does not account for changes in background population distribution [39]. However, as this study uses data collected over a brief—30 week—period, population sizes in each county are not

expected to have changed significantly, and population density was included as an additional covariate.

In conclusion, spatiotemporal methods applied to reported COVID-19 deaths data identified clusters of mortality that are not explainable by case patterns alone. Other known drivers of disease severity at the individual level, including race and age, are significantly associated with county-level deaths. Together with other county covariates and environmental variables, these appear to explain patterns of both lower or higher mortality rates than expected, as models controlling for these covariates do not identify any remaining clustering in space or time.

Table 2.1: Characteristics of counties within and outside of reported clusters:

	Counties included in high mortality clusters (N=120)	Counties included in low mortality clusters (N=1641)	Counties not included in a cluster (N=1295)
Population density			
Mean (SD)	1040 (1940)	94.2 (270)	229 (825)
Median [Min, Max]	403 [4.6, 13900]	34.4 [0.3, 5200]	57.9 [0.3, 17600]
Proportion Black			
Mean (SD)	0.078 (0.074)	0.106 (0.158)	0.079 (0.128)
Median [Min, Max]	0.057 [0.005, 0.419]	0.028 [0, 0.866]	0.021 [0, 0.804]
Proportion Hispanic			
Mean (SD)	0.136 (0.149)	0.072 (0.090)	0.127 (0.179)
Median [Min, Max]	0.068 [0.013, 0.850]	0.039 [0.006, 0.742]	0.054 [0.007, 0.964]
Proportion 65+			
Mean (SD)	0.188 (0.042)	0.201 (0.044)	0.196 (0.051)
Median [Min, Max]	0.182 [0.112, 0.398]	0.199 [0.071, 0.375]	0.188 [0.049, 0.582]
Proportion food insecure			
Mean (SD)	0.106 (0.029)	0.136 (0.04)	0.129 (0.034)
Median [Min, Max]	0.103 [0.052, 0.192]	0.134 [0.036, 0.304]	0.128 [0.038, 0.270]
County's own ICU beds			
Mean (SD)	90.2 (219)	7.81 (29.2)	19.7 (64.2)
Median [Min, Max]	21.5 [0, 2000]	0 [0, 469]	0 [0, 1030]
Neighboring county's ICU beds			
Mean (SD)	85.9 (112)	8.39 (14.2)	20.3 (32.4)
Median [Min, Max]	54.4 [4.38, 670]	3.71 [0, 164]	8.00 [0, 303]

Table 2.2: Details of clustering timing, observed deaths, and expected deaths.

Cluster	Lower or higher mortality	Time frame	Observed deaths	Expected deaths
1	Lower	March 2 - April 6	903	2011.97
2	Lower	March 2 - April 6	311	915.71
3	Higher	March 2 - March 30	2286	1474.85
4	Higher	March 2 - March 30	291	91.34
5	Lower	March 2 - April 13	25	149.03
6	Higher	March 16 - March 23	114	26.64
7	Higher	March 16 - April 6	151	55.97
8	Higher	March 9 - March 23	76	20.39

Table 2.3: Negative binomial model (log link) results.

Variable	Estimate	exp(Estimate)	p-value
Temperature (1 week lag)	-0.249	0.7796	< 0.0001
Relative Humidity (1 week lag)	-0.02	0.9802	0.0339
Population Density	0.205	1.2275	< 0.0001
Proportion 65+	0.022	1.0222	0.0275
Proportion Black	0.205	1.2275	< 0.0001
Proportion Hispanic	0.152	1.1642	< 0.0001
Proportion Food Insecure	-0.162	0.8504	< 0.0001
Base County's ICU Beds	0.106	1.1118	< 0.0001
Neighboring County's ICU Beds	-0.002	0.9980	0.8264

CHAPTER 3

COMMUNITY DRIVERS OF TUBERCULOSIS DIAGNOSTIC DELAY IN KAMPALA, UGANDA: A RETROSPECTIVE COHORT STUDY[2]

[2] Mercaldo, R.A., Whalen, C.C., Kakaire, R., Nakkonde, D., Handel, A., Sekandi, J.N. Submitted to *BMC Infectious Diseases*, 2/9/2021

Abstract

Recent approaches to TB control have focused on identifying and treating active cases to halt further transmission. Patients with TB symptoms often delay to seek care, get appropriate diagnosis, and initiate effective treatment. These delays are partly influenced by whom the patients contact within their community network. We aimed to evaluate the community drivers of diagnostic delay in an urban setting in Uganda.

In this study we analyze data from a retrospective cohort of 194 TB patients in Kampala, Uganda. We characterized the patterns of contacts made by patients seeking care for TB symptoms. The main outcome of interest was total community contact delay, defined as the time patients spent seeking care before visiting a provider capable of diagnosing TB.

Visits to health providers without access to appropriate diagnostic services accounted for 56% of contacts made by cohort members, and were significantly associated with community contact delay, as were symptoms common to other prevalent illnesses, such as bone and joint pain. Education programs aimed at primary care providers, as well as other community members, may benefit case identification, by informing them of rarer symptoms of TB, potential for coinfections of TB and other prevalent diseases, and the availability of diagnostic services.

Introduction

Tuberculosis is one of the top ten leading causes of death worldwide, and the leading cause of death by a single infectious agent in 2019 [41]. While TB prevalence studies indicate that infected individuals may transmit tuberculosis bacilli before symptom onset [42], the majority of transmission occurs between the debut of symptoms and treatment initiation [43].

Contagiousness, as measured by bacillary numbers on sputum smears, increases with treatment delays [44]. In the absence of a broadly effective vaccine, control measures rely on shortening this transmission period through early diagnosis and treatment of active pulmonary disease [43,45].

Globally, the ideal of prompt identification and treatment of TB is not yet a realized norm; in many settings there is a fraction of patients who only receive diagnosis and treatment after a prolonged delay [46]. An extensive body of literature is dedicated to studies of delays at both the patient and healthcare system levels, in a variety of incidence or socio-economic settings. These studies have reported numerous risk factors for delay, including comorbidities [47,48], low access to healthcare [49–51], initial visits to low-level healthcare facilities with inadequate diagnostic abilities [52–54], age and sex [48,55], and beliefs or misunderstandings about the disease [56–58]. These factors affect delay in one or more stages by increasing the duration of time spent 1) experiencing symptoms without seeking care, 2) searching for qualified practitioners, or 3) awaiting diagnosis following a visit to a qualified provider or facility.

Categories of diagnostic delay have been summarized as either *health system delay* or *patient delay* [59]. Here, *health system delay* refers to the time from first contacting a qualified TB provider to final diagnosis and treatment, and may be related to numerous factors including sex of the patient [43,60] or symptom profiles [61–63]. *Patient delay* refers to both the individual's delay in seeking care and the time spent contacting unqualified providers or social contacts. While many studies have combined these patient delay portions of the diagnostic pathway for analysis, they are two distinct periods in which behavior and other drivers of delay likely differ. In our present study, we divide patient delay into its two component periods. *Care-seeking delay* was defined as a participant's symptomatic time prior to seeking care. We defined

community contact delay as the time spent actively seeking care in the community. Within this latter period, an individual may seek advice or help from any member of their community. This could include social contacts such as family or workmates, hereby termed social contacts, non-TB providers such as primary-level health providers or herbal or religious healers.

We maintained the definition of *healthcare delay* as the time from a participant's first contact with a qualified TB provider to the point of final diagnosis. The different types of periods are summarized and illustrated in Figure 3.1.

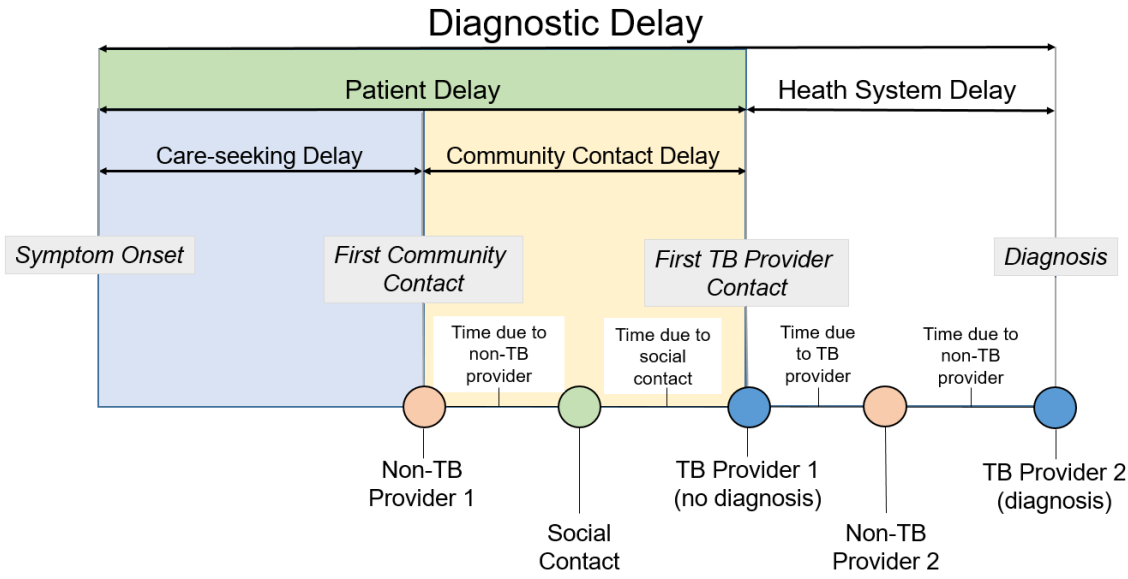

Figure 3.1: Conceptual Framework, with study definitions and an example of a realized diagnostic pathway.

Previously, we conducted a retrospective cohort study (Diagnostic I) to quantify diagnostic delay in Kampala, Uganda, focusing on TB patients as members of a broader community of social contacts and health providers [45]. We examined patterns in delays to diagnosis among TB patients in urban health clinics, quantifying care-seeking delays and collecting detailed information regarding the social and provider contacts that participants make upon initiation of care-seeking. In a follow up study, Diagnostic II, we have extended findings of

the first study to examine additional factors contributing specifically to delay in the community. While other studies have focused on factors associated with decisions to seek care [64,65] or on health system delays [43,51,61,66], the present study examines factors associated with increased or decreased delay in the *community-contact* portion of the diagnostic pathway.

Methods

We analyzed data from Diagnostic II, the second of two retrospective cohort studies conducted in Kampala, Uganda. This second study expanded on the methods of Diagnostic I, described previously [45].

Ethical considerations

Written informed consent was obtained from all eligible participants. The study was approved by institutional review boards at the University of Georgia, Makerere University School of Public Health, and the Uganda National Council for Science and Technology.

Study design, setting, and population

We conducted a retrospective cohort study among TB patients from January to November 2017. Participants were recruited at two public TB clinics located in Lubaga Division, and within 5-10 km of Kampala, Uganda's capital city. The clinics are part of the government-funded public health system run by the Kampala Capital City Authority. Primary health care services, diagnosis and treatment of TB and other health conditions are provided free of charge. The estimated catchment population of the public clinics in Lubaga division is 400,000 persons. Additional health facility census information for the study area in 2017 is available from the United States Agency for International Development [67]. Eligible patients were consenting adults, eighteen years or older, who had been diagnosed with active pulmonary tuberculosis and who had initiated treatment within three months of the interview date. Participants were recruited

at variable times after diagnosis and were interviewed to collect retrospective information on time of seeking care before diagnosis; this approach was previously deemed a suitable alternative to prospective cohort studies [68].

Data collection and management

Data were collected in face-to-face interviews by trained interviewers using a structured questionnaire (available via our Github repository). The questionnaire was developed by a team of physicians, with expertise in TB, and epidemiologists. The original questionnaire used in our first study, Diagnostic I, was tested in a pilot study for accuracy, comprehension, and consistency of responses, with satisfactory results [45]. For Diagnostic II, the questionnaire was expanded to include items about participant knowledge about TB symptoms, experiences with and concerns about TB symptoms, prompts to seek care, and costs of reaching or obtaining health care. These variables were additions to the original items on HIV status, time of TB diagnosis, time of onset of symptoms, and duration of symptoms, as well as the detailed information about contacts made while seeking care. The complete list of variables is included as supplemental material.
Data were collected using standardized teleforms and scanned into a database using optical scanning software (TeleForms®). We preprocessed the raw data and engineered summary or comprehensive factors relevant to the analysis when applicable. All numeric variables were centered and scaled. All code and additional details are available via our Github repository.

Descriptive analysis

We calculated community contact delay as the time from first seeking care to first contacting a qualified TB provider. Qualified TB providers included government hospitals, government health centers, private hospitals, or other locations with TB diagnostic services.

For the analysis of these community delays, contacts were divided into two categories: social contacts and non-TB providers. Social contacts included spouses, parents, children, siblings, other relatives, coworkers, friends, and neighbors. Non-TB providers included herbal healers, drug stores, private clinics, or village health workers. The time contributed to a patient's pathway was decomposed into steps between contacts, and each window of time was considered related to the most recent contact. In this way, the total community contact delay could be divided into the times specific to visits to contacts in each category. We calculated additional measures including the number and fraction of community network contacts in each category, as well as the total number of contacts and the total amount of time spent visiting contacts.

The outcome of interest was total community contact delay. As visits to non-TB providers were significant in the Diagnostic I study [45], a secondary analysis was included to explore factors associated with the number of community contact delay days contributed by visits to non-TB providers.

Statistical analysis

We fit linear regression models with each predictor individually, to investigate bivariate associations with community contact delay. Similarly, we fit bivariate regression models for each predictor for our secondary analysis, investigating the contribution of time spent contacting non-TB providers to overall community contact delay.

Two final linear models were fit with Least Absolute Shrinkage and Selection Operator (LASSO) regularization and 10-fold cross validation—one each for the outcomes of (1) community contact delay and (2) the contribution of non-TB provider visits to community contact delay. The distribution of the residuals for full linear models with all predictors showed some skewness. Neither a log-transformation of the outcome nor use of Poisson distribution

models improved the minor skew, and linear regression was maintained for the final LASSO models. All analyses were conducted in R software (version 3.6.1) [33].

Results

Table 1 reports the characteristics of the Diagnostic II study population. Of the 194 study participants in Diagnostic II, 177 (91.2%) were new TB patients, while only 17 (8.8%) were retreatment cases. The mean age of participants was 32 years (sd: 11.7 years), and 62.4% were male. There were 63 (32.5%) HIV positive participants and 129 (66.5%) who were HIV negative (Table 1).

Patterns of community contact delay

The Diagnostic II population spent a total of 9014.69 days in the community contact period, with a median 33 days (IQR: 14-66.75) spent by each patient. The 194 participants reported visiting 895 contacts during this period. Of these, 397 (44.4%) were social contacts, while 498 (55.6%) were non-TB providers. Patients made a median 5 contacts within their community before reaching a qualified provider. Though the Diagnostic II cohort made approximately 25% more contacts with non-TB providers, the actual time contributed to overall community contact delay by visits to such providers (4625.56 days), was similar to that contributed by visits to social contacts (4378.13 days) (Figure 3.2).

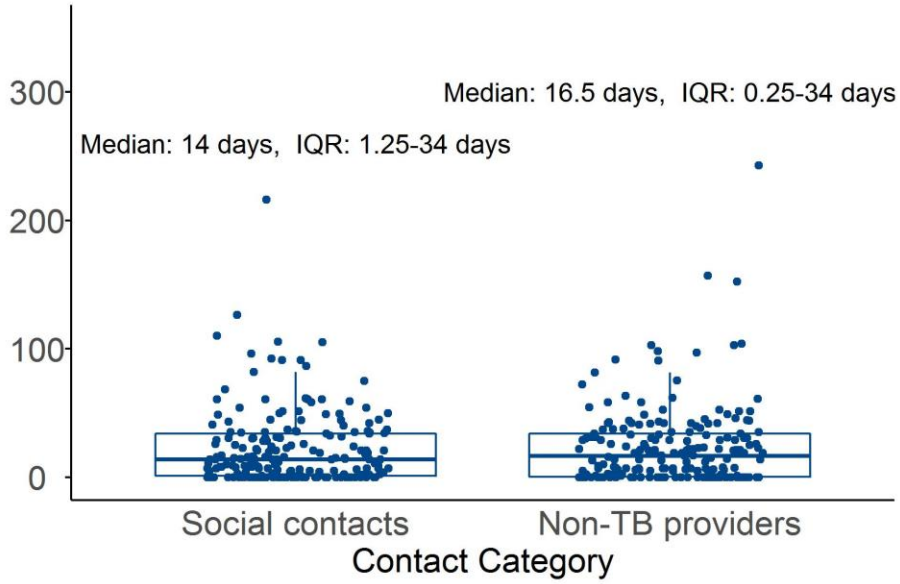

Figure 3.2 Community Contact Delay. Time (in days) contributed to community contact delay by visits to social contacts or primary-level, non-TB providers in the II study. Each point represents one patient, with median and IQR shown.

Model results – community contact delay

Bivariate regression models of community contact delay were fit for all available predictors. As there were 49 predictors, the table of all results is too large to include in text and is available as supplemental material. Table 3.32 reports the results for the twelve significant variables ($p < 0.05$). A 10% increase in the proportion of contacts made to non-TB providers (rather than social contacts) was associated with an additional 1.2 days of delay in the community. Receiving cough medication was associated with over 39.3 days of additional delay ($p = 0.001$), while each additional receipt of cough medication was associated with 11.4 days of delay ($p = 0.0005$). Suspicion that the illness was TB was associated with 16.4 fewer days of delay ($p = 0.0143$). Specific symptom knowledge or experiences were also associated with decreased delay: knowing that appetite loss or weight loss is a symptom of TB (18.4 fewer delay days, $p = 0.0197$), knowing that coughing blood or chest pain is a symptom of TB (14.2 fewer delay days, 0.0306), or seeking evaluation for TB due to night sweats or fever (15.2 fewer delay

days, 0.0217)). Conversely, experiencing or feeling concern over bone or joint pain symptoms, or seeking care for bone or joint pain symptoms, was associated with increased delay (21.3 delay days and 39.3 delay days, respectively, $p = 0.0032$, $p = 0.0118$).

In a final linear model with LASSO regularization and 10-fold cross-validation, fourteen variables were selected and are reported in Table 3.3. Experiencing or feeling concern about bone or joint pain was associated with 16.8 additional delay days. Seeking care for coughing blood or chest pain was also associated with an increase of 0.7 delay days, though the participant knowing these are symptoms of TB was associated with decreased delay (2.7 days). Evaluation for TB due to night sweats or fever was associated with 7.8 fewer delay days, while finding no relief from self-medicating was associated with 3.4 fewer delay days. Receiving cough medication was associated with 21.5 additional days of delay. Notably, a 10% increase in the proportion of contacts in the non-TB provider category was associated with only 0.58 additional days of delay in the final model.

Model results – contribution of visits to non-TB providers to community contact delay.

As visits to contacts in the non-TB provider category was found significant in our original study, Diagnostic I [45], we also analyzed factors associated with the time visits to these contacts contributed to overall community contact delay in the present study. In bivariate analysis (Table 3.4), receipt of cough medication was associated with 23.3 days of delay following visits to these providers, ($p = 0.0016$). Each time medication was received was associated with 8.9 delay days ($p = 0.0001$). Increases in total cost of reaching care (in terms of Ugandan shillings, centered and scaled) was significantly associated with increased delay (6.56 days, $p = 0.0036$). Symptom knowledge was associated with decreased delay: 10.8 fewer days for knowing appetite loss or weight loss is a symptom of TB ($p = 0.0481$) and 9.7 fewer days for

knowing coughing blood or chest pain was a symptom of TB ($p = 0.0313$). Seeking care for bone or joint pain was associated with 37.4 additional delay days ($p = 0.0004$), while being evaluated for TB due to bone or joint pain was associated with 21.3 additional days of delay following visits to non-TB providers ($p = 0.0148$).

In a final linear model with LASSO regularization and 10-fold cross-validation, sixteen variables were selected and are reported in Table 3.5. A number of these factors are related to symptoms: experiencing or being concerned about cough-related symptoms or malaise were associated with 3.6 and 4 fewer days of delay, respectively, following visits to non-TB providers. Seeking care for bone and joint pain was associated with 22 additional delay days, while seeking care for cough symptoms was associated with 0.7 additional days. Symptom knowledge and the participant's suspicion that the illness was TB were once more associated with decreasing delays, and receipt of cough medication again associated with increasing delays (Table 3.5).

Discussion

In this exploratory analysis of the Diagnostic II retrospective cohort data, we found that tuberculosis patients sought care from their community contacts for a median of 33 days before contacting a health care professional at a government-designated TB service clinic. In total, the 194 participants spent 9,015 cumulative days actively seeking care, indicating that substantial delay occurs during the community contact portion of the diagnostic pathway.

Ideally, patients should seek care from specialized TB providers upon recognizing symptoms. In the Diagnostic I study, we suggested that patients, possibly unaware of the cause of their illness, may first seek care through another care provider or social contact who then refers them to the appropriate diagnostic location. In the present study, we found that patients made a median 5 such contacts before reaching a qualified provider.

Previous studies have shown that patients cycle through repeated visits to lower-level, primary, health providers [43,66,69], and our Diagnostic I study shows a significant portion of diagnostic delay, within the community contact portion of the pathway, was spent in visits to non-TB providers. We have recommended interventions targeting non-TB providers to reduce these delays [45]. In the present study, we show that, while the actual delay time due to non-TB providers was similar to that of social contacts, patients made 25% more visits to non-TB providers.

In a secondary analysis, we focused on the time contributed to overall community contact delay by visits to non-TB providers within the Diagnostic II population. We found that a symptom attributable to other prevalent febrile diseases, bone and joint pain, was significantly associated with the time spent in visits to non-TB providers. Further, bone and joint pain symptoms, as well as cough-related symptoms, were selected in a linear model with LASSO regularization. With these results, similar to those found in other studies [43,61], we maintain the recommendation that non-TB providers complete continuing education emphasizing TB screening, even in such cases when malaria, typhoid, or other febrile illnesses are suspected. Additionally, they should be encouraged to refer patients to proper diagnostic locations or recommend them for follow-up, possibly through active case finding.

We acknowledge that coinfection of TB and other illnesses that cause fever, such as HIV or malaria, is certainly a possibility. The present study is not able to identify such cases, or the cause of symptoms common to other diseases (such as bone and joint pain in malaria). Further research on the prevalence of coinfection and associated symptom profiles would improve education efforts.

We assumed that each contact in a patient's diagnostic pathway was a separate, non-overlapping event, and so our calculated community contact delay may be overestimated. Additionally,

our data collection relied on patient-reported details, such as the length of time they experienced symptoms or the time between contacts, and as such is subject to recall bias. We recruited participants who had been diagnosed with TB within only three months of their interview, to decrease bias in this area.

While many studies have analyzed data on patient delays to diagnosis, our analysis focuses on a unique period in the diagnostic pathway—the time spent seeking care and contacting members of the community until a final diagnosis is reached. Recommendations for shortening delay at this stage may differ from those made to shorten care-seeking delays (when symptomatic patients have not yet begun to seek care). Community-based TB programs often focus on recognition of common TB symptoms—chronic cough, weight loss, night sweats, and fever—and encourage those with symptoms to visit health facilities or otherwise seek care. Our results suggest that further delays, once the patient is engaged in seeking care, may depend on interactions in the community and, particularly, with lower-level healthcare providers. Education efforts targeted for specific audiences (non-TB providers versus social contacts or the patients themselves) might focus on rarer symptoms of TB, or the wisdom of visiting TB diagnostic locations despite recognizing symptoms more common to other prevalent diseases. Some models of creating mass awareness about TB have been proposed and used elsewhere [70]. To facilitate appropriate actions that shorten community contact delays, improved point of care (POC) diagnostic TB tests that are delivered at the most decentralized levels of care where the patients make the initial contact with the non-TB provider health system, as well as within the

community, are needed [71]. The use of POC at community level would minimize any barriers or further delays in case detection that are introduced during the referral process to TB service centers.

In conclusion, the Diagnostic II cohort spent 9015 days seeking care, indicating a substantial portion of diagnostic delay occurs while TB patients contact members of their community. Seeking care from primary-level providers is associated with increased delay in the community, and symptoms common to prevalent febrile illnesses significantly increases the time spent visiting such providers. Improving point of care (POC) diagnostics within local communities and continuing education for primary-level providers may benefit case identification and decrease overall diagnostic delays.

.

Table 3.1: Baseline characteristics of participants in the Diagnostic II study.

	(n = 194)
Sex	
Female	72 (37.1%)
Male	121 (62.4%)
Missing	1 (0.5%)
Age (Years)	
Mean (SD)	32.0 (11.7)
Median [Min, Max]	28 [18, 82]
Missing	2 (1%)
Marital Status	
Currently married	69 (35.6%)
Not married	125 (64.4%)
Monthly Income (UGX*)	
Mean (SD)	294,000 (481,000)
Median [Min, Max]	200,000 [0, 5,000,000]
TB Episode	
New case	177 (91.2%)
Retreatment	17 (8.8%)
HIV Status	
Negative	129 (66.5%)
Positive	63 (32.5%)
Unknown	2 (1%)

**UGX, Ugandan Shillings. In the study year, 2017, the conversion rate for 1 US dollar was 3616.24 UGX.*

Table 3.2: Results of bivariate linear regression for significant ($p < 0.05$) predictors in models of community contact delay

Variable	Estimate	Std. Error	Pr(>|t|)
Proportion of contacts in non-TB provider category (10% increments)*	1.197	3.198	0.0002
Number of times cough medication received	11.351	3.202	0.0005
Total cost for care**	10.873	3.219	0.0009
Suspected illness was TB	-16.363	6.616	0.0143
Received cough medication	40.610	10.465	0.0001
Cough disrupted daytime activity	17.978	6.598	0.0070
Knows appetite loss or weight loss is symptom of TB	-18.447	7.844	0.0197
Knows coughing blood or chest pain is symptom of TB	-14.233	6.532	0.0306
Experienced, or was concerned about, bone or joint pain	21.316	7.129	0.0032
Sought care for bone or joint pain	39.271	15.454	0.0118
Evaluated for TB due to bone or joint pain	25.671	12.636	0.0436
Evaluated for TB due to night sweats or fever	-15.150	6.547	0.0217

**Coefficient should be read as the increase in delay days associated with each increase of 0.1 in the proportion of contacts that belong in the non-TB provider category*

***UGX, Ugandan Shillings. In the study year, 2017, the conversion rate for 1 US dollar was 3616.24 UGX.*

Table 3.3: Results of linear regression (with LASSO regularization) models of community contact delay

	Model R^2: 0.29
Variable	**Estimate**
Experienced, or was concerned about, bone or joint pain	16.755
Sought care for coughing blood or chest pain	0.659
Evaluation for TB due to night sweats or fever	-7.808
No relief from self-medication prompted care-seeking	-3.368
TV/Radio advertisement prompted care-seeking	5.223
Knows coughing blood or chest pain is a symptom of TB	-2.730
Knows appetite loss or weight loss is a symptom of TB	-4.945
Received cough medication	21.485
Cough disrupted daytime activity	3.880
Suspected illness was TB	-8.955
Bought supplements	-0.516
Diagnosis location - outside Rubaga	-0.293
Proportion of contacts in non-TB provider category (10% increments)*	5.763
Total cost of reaching care**	4.448

*Coefficient should be read as the increase in delay days associated with each increase of 0.1 in the proportion of contacts that belong in the non-TB provider category

***UGX, Ugandan Shillings. In the study year, 2017, the conversion rate for 1 US dollar was 3616.24 UGX.*

Table 3.4: Results of bivariate linear regression for significant ($p < 0.05$) predictors in models of contribution of non-TB provider visits to community contact delay

Variable	Estimate	Std. Error	Pr(>|t|)
Number of times cough medication received	8.851	2.197	0.0001
Total cost for care*	6.559	2.228	0.0036
Suspected illness was TB	-9.888	4.657	0.0350
Received cough medication	23.326	7.278	0.0016
Knows appetite loss or weight loss is symptom of TB	-10.768	5.413	0.0481
Knows coughing blood or chest pain is symptom of TB	-9.739	4.490	0.0313
Sought care for bone or joint pain	37.377	10.457	0.0004
Evaluated for TB due to bone or joint pain	21.248	8.643	0.0148

**UGX, Ugandan Shillings. In the study year, 2017, the conversion rate for 1 US dollar was 3616.24 UGX.*

Table 3.5: Results of linear regression (with LASSO regularization) models of contribution of non-TB provider visits to community contact delay

	Model R^2: 0.27
Variable	**Estimate**
Experienced, or was concerned about, coughing blood or chest pain	-3.628
Experienced, or was concerned about, malaise	-3.964
Sought care for bone or joint pain	21.984
Sought care for coughing blood or chest pain	0.733
Evaluation for TB due to night sweats or fever	-0.528
Knows coughing blood or chest pain is a symptom of TB	-3.322
Knows appetite loss or weight loss is a symptom of TB	-0.779
Received cough medication	5.033
Suspected illness was TB	-4.101
Someone other than participant expressed concern about symptoms	-1.182
Bought supplements	-4.512
Age (years)	0.044
Marital status - currently married/cohabiting	-0.064
TB episode - first episode	-0.052
Total cost of reaching care*	7.850

**UGX, Ugandan Shillings. In the study year, 2017, the conversion rate for 1 US dollar was 3616.24 UGX.*

CHAPTER 4

EVALUATION OF ALTERNATIVE ENDPOINTS FOR ZIKV VACCINE EFFICACY TRIALS[3]

[3] Mercaldo, R.A., and Bellan, S.E. *Vaccine*. 2019; 37(15):2099-2105.
Reprinted here with permission of the publisher.

Abstract

Zika virus (ZIKV) infection during pregnancy is associated with microcephaly and other birth defects, collectively termed Congenital Zika Syndrome (CZS). During the epidemic in 2015-2016, ZIKV spread through the Americas and quickly joined the list of other known teratogenic pathogens, TORCH. Multiple ZIKV vaccines have been developed for protection of pregnant women and women of childbearing age. However, ZIKV infection incidence has since waned substantially, and adverse birth outcomes are rare outcomes of infection. Studying a vaccine's protective efficacy against CZS in a large phase III clinical trial may be infeasible in such times of low incidence. Should trials be initiated, researchers may resort to alternative clinical endpoints.

In this study, we simulate a variety of vaccine clinical trial scenarios to evaluate the feasibility of the CZS endpoint in vaccine studies and compare CZS to other potential outcomes: ZIKV infection detected through weekly, biweekly, or monthly testing and laboratory-confirmed, symptomatic Zika Virus Disease. We compare the sample size required for 80% statistical power to detect vaccine efficacy and trial duration for each scenario. Our results show the feasibility of CZS clinical endpoints depends on the timing of simulated clinical trials in the course of a seasonal epidemic, due to CZS risk varying with trimester of infection. This result highlights additional considerations needed when designing vaccine efficacy trials of protection against teratogenic pathogens.

Introduction

Zika virus (ZIKV) was first isolated in Uganda in 1947 and associated with human illness in 1953 [72]. Sporadic outbreaks followed until 2015, when the virus was identified in Brazil and linked to an alarming increase in the incidence of microcephaly and other adverse birth outcomes [72–75]. Researchers began to list ZIKV with other infectious teratogens, exposures that harm a developing embryo or fetus, which include pathogens known by the acronym TORCH—Toxoplasmosis, other agents, Rubella, Cytomegalovirus, and Herpes Simplex [76].

ZIKV spread to over fifty countries and territories after its introduction to the Americas [77]. During that time, its teratogenic effects led the World Health Organization (WHO) to declare a Public Health Emergency of International Concern and call for the development of vaccines to protect pregnant women and women of childbearing age [78]. WHO removed the emergency designation in November 2016, but ZIKV remains a significant threat during pregnancy, with an extensive list of potential adverse birth outcomes collectively termed Congenital Zika Syndrome (CZS) [79]. In response, multiple ZIKV vaccines were developed and entered early phase clinical trials. Before these vaccines are made available to the public, large phase III trials must be completed to evaluate their efficacy, the reduction in risk in the vaccinated compared to the unvaccinated [80].

As a phase III clinical trial's power to detect vaccine efficacy depends in part on the number of cases observed, a successful trial will depend on the extent of ZIKV transmission in the study area. While ZIKV incidence has fortunately waned substantially since the initial outbreak, this poses challenges to the successful identification of a safe and efficacious vaccine. The Zika Modeling and Projections for Vaccination Trials collaboration recommended increasing the number of study sites in various geographical regions, along with recruiting larger

numbers of participants and extending the duration of trial participant follow-up to adequately power a vaccine efficacy trial [81]. Furthermore, though WHO called for vaccines that protect pregnant women and women of childbearing age, CZS is a rare outcome of ZIKV infection and may not be a feasible clinical trial endpoint in a time of low incidence [82]. Other possible endpoints are ZIKV infection and symptomatic Zika Virus Disease, both of which occur at higher frequency and would allow recruitment from a larger population of both men and women.

Choosing either infection or symptomatic disease as primary endpoints in a ZIKV vaccine trial can only be justified, however, if protection against them reasonably predicts protection against more severe complications, like CZS, that are the greatest health burdens associated with the virus [82]. This is challenging for symptomatic disease endpoints, as disease has not yet been associated with higher viral load or increased risk of maternal-fetal transmission [83]. Conversely, vaccines that prevent maternal infection may be expected to prevent subsequent maternal-fetal transmission, and research suggests that complete sterilizing immunity may be required to protect against CZS [84]. In practice, testing for ZIKV infection relies on viral RNA detection [82,85,86], and the tests used have a lower limit of detection that may be above the viral load necessary for maternal-fetal transmissions [84,87]. In this scenario, trials can show efficacy against *detectable viral load*, when infection may still have occurred and sterilizing immunity was not produced. For simplicity, we continue to refer to this clinical endpoint as infection-based, though we recognize that it may be impossible to show actual sterilizing immunity.

It is clear that the choice of clinical endpoint has far-reaching consequences for the feasibility of a ZIKV vaccine trial and the ultimate conclusions about the vaccine's protective efficacy against the most severe outcomes of infection. These consequences include immediate

effects on trial participant recruitment and statistical power to detect vaccine efficacy. We simulate a variety of clinical trial scenarios in the present study, to identify important patterns to consider when evaluating the feasibility of the following ZIKV clinical endpoints: ZIKV infection detected through weekly, biweekly, or monthly testing, laboratory-confirmed Zika Virus Disease, and CZS.

Methods

We simulated ZIKV vaccine efficacy trials to explore three main clinical endpoints: infection, symptomatic disease, and CZS. Within infection-based endpoints, we additionally varied the time between laboratory tests. In all cases, simulations were completed in four steps. First, we obtained observed ZIKV incidence data for six countries and scaled the data to meet a generic 1% cumulative annual incidence rate while maintaining the seasonal pattern observed. Second, we simulated trial populations of various sizes within those regions, and assigned weekly risk according to seasonal patterns, layering individual heterogeneity in infection risk on top of these patterns. Third, we modeled time until infection, time until symptom onset, and time until loss of detectable viral RNA in blood samples for a range of assumed vaccine efficacies in trials with ZIKV infection and symptomatic Zika Virus Disease endpoints. For CZS endpoint trials, we additionally generated CZS outcomes diagnosed at birth (see Table 4.1 for simulation parameters). Finally, we analyzed simulated trial data to assess the statistical power of each trial scenario, and compared sample size and trial duration for trials achieving 80% power. We used R version 3.4.2 for all simulations and analyses [33].

Scaling weekly country-level incidence rates

The Pan American Health Organization publishes periodic updates on ZIKV cases in Latin America and the Caribbean. These updates include tables of cumulative case counts and

are often accompanied by graphs of the ZIKV epidemic curve in each country [77]. There are notable errors in the case count tables, which are fortunately not reflected in the graphs of epidemic curves [88]. Rather than use the tabular data and risk including these errors in our simulation, we obtained the graph data of weekly incidence in multiple countries in 2016 and 2017 that was previously digitized by the Andersen Lab at the Scripps Research Institute [88]. For this study, we obtained the digitized data for eight countries previously identified as having sufficient infrastructure for a large clinical trial [81]. Of these, six countries had eighty-two weeks of observed incidence data at the time of our study: Colombia, Costa Rica, Ecuador, Mexico, Panama, and Peru. The two remaining countries were excluded for having missing observations (Brazil) or many weeks with 0 cases (Dominican Republic), which made the data unscalable. We scaled the digitized incidence data to meet a generic national yearly cumulative incidence rate, maintaining the seasonal pattern of Zika transmission in our simulation (Fig. 4.1).

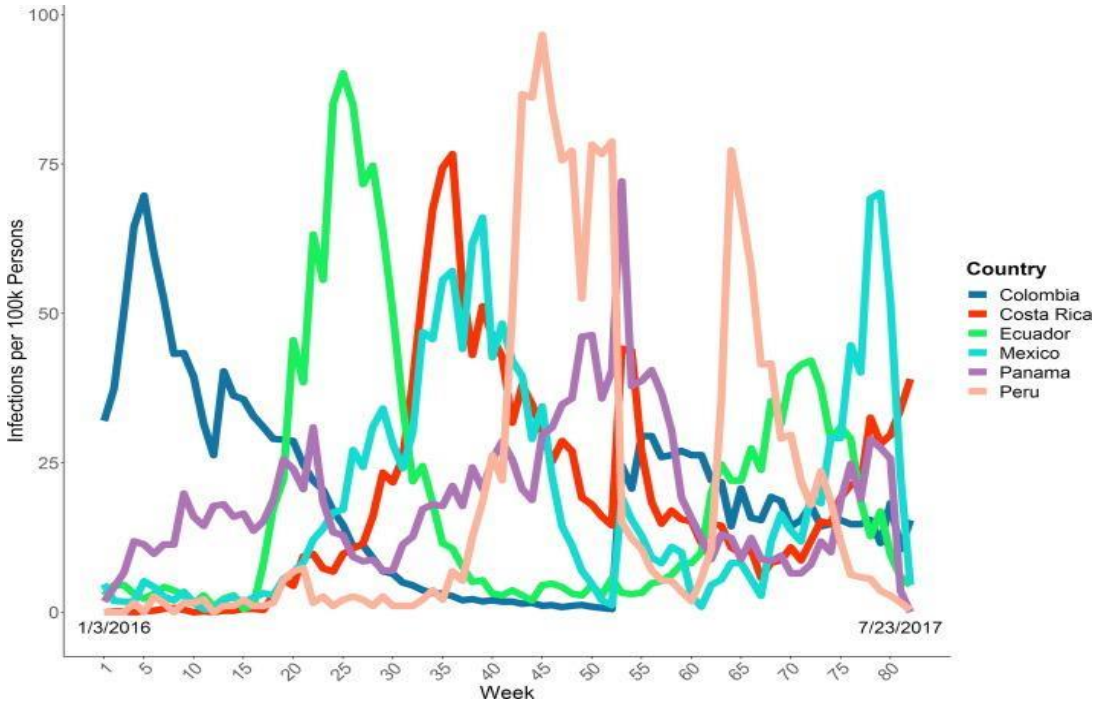

Figure 4.1. Scaled person-week infection rates for the six countries included in the study.

A generic cumulative incidence rate of 1% of the population per year was chosen, not to reflect actual epidemic values but to better illustrate qualitative patterns between trial endpoints. To imitate realistic trial site selection, we additionally modeled a selection process so that four countries would ultimately be chosen, and that selection would be based upon monitored ZIKV infection incidence. To simulate this process, we randomly selected four of the six countries at the beginning of each simulation iteration. To examine how incidence at the start of a trial affected feasibility, we set trial start dates independently within each of the four countries according to three possible starting rules: (1) when that country first reached a weekly ZIKV infection rate of 25 cases per 100,000 persons in the scaled data, (2) when it reached a weekly rate of 50 cases per 100,000 persons, or (3) one week following the peak of a local epidemic in the first 40 weeks (approximately half) of the scaled data.

Simulating trial populations

Each participant's weekly ZIKV infection risk was modeled as the product of the country-level weekly incidence rate and an individual-level risk factor. Individual-level infection risk factors were drawn from the lognormal distribution (log mean = 0, log sd = 1) to simulate heterogeneity in infection risk that was non-negative. We assumed independence in infection risk, as trial populations are likely to be a small proportion of each country's total population if participants are recruited from many countries. Though only six countries were used in this study, recruiting from a large number of total countries is recommended for ZIKV vaccine trials [82].

In each simulation, we varied the start time of the trial according to one of three possible starting rules, discussed above. Each trial participant was then randomized to either the control or vaccination arm. We considered vaccine efficacies of 50%, 70%, and 90%, and assumed an

immune ramp-up period lasting one month before achieving the assumed efficacy. Efficacy was simulated as a reduction in each vaccinated individual's overall weekly risk. We simulated a point recruitment process such that all individuals were randomized and, if applicable, vaccinated at the trial start date.

Generating clinical outcomes

We generated time until infection for each participant using the exponential distribution, so that a participant's risk of infection in a given week was independent of previous weeks. We assumed viral RNA would be detectable in blood two days after infection. We then used reported estimates for the mean time to viral clearance to model an RNA detection interval using the flexible Weibull distribution [89,90]. The detection interval described the time from first positive samples until the virus could no longer be detected in blood.

For trials with ZIKV infection as the primary endpoint, we created a set of laboratory testing days that began at the start of the trial and occurred weekly, biweekly, or monthly thereafter. Cases were identified when these predetermined testing days fell within an individual's RNA detection interval. In simulated trials with laboratory-confirmed ZVD as the primary endpoint, we determined which infected individuals would develop disease based on the approximate 20–25% symptomatic rate [82]. A median incubation period of 5.9 days was used to generate time to symptom onset [89]. Viral RNA was assumed to be detectable upon symptom onset. The duration of detection time in samples was generated as above. Rather than record all ZVD outcomes as events, cases were identified when weekly testing dates fell within a symptomatic individual's detection interval. Weekly testing was used as it was assumed not all individuals would report to a study clinic immediately, but may be expected to do so within a short timeframe given an intense surveillance protocol as suggested by WHO [82]. For those

cases identified by this protocol, we recorded time until symptom onset as time until infection plus the incubation period.

For trial simulations in which CZS was the primary endpoint, we assumed an entirely female study population. Furthermore, we focused on populations of women who self-identify as trying to conceive (TTC). Women who first attempt to conceive have a 30% probability of conception within a menstrual cycle, and this probability decreases to approximately 5% in women who have not conceived by the end of the first year [91]. In the CZS endpoint scenarios, all the women in the trial were given an initial 21.3% probability of conception in the first month of the trial. This probability was calculated as the mean of the first six months of exponentially decaying conception probability, to account for women entering the trial with different histories of conception attempts, and decreased every four weeks of the trial. Time until pregnancy was drawn from the exponential distribution. We set CZS risk according to trimester of ZIKV infection: infection in the first trimester only (15%), infection in multiple trimesters, including the first (21%), or infection in the second or third trimesters only (2.27%) [5]. The CZS outcome was considered binomial, as either present or absent at birth, and so was generated from the binomial distribution.

In sensitivity analyses, we included women regardless of desire to conceive. Given the contraceptive use rate of 73% in Latin America [92], we used the complement of this rate as the assumed proportion of trial participants who were TTC. While this most likely overestimates the proportion of women who are sexually active and actually attempting to conceive, none of these trials achieved adequate statistical power in simulations and were not examined further.

We varied trial population sizes from 200 to 30,000 individuals for non-CZS endpoints and from 200 to 120,000 individuals for CZS endpoints. We simulated 250 runs for each scenario (i.e. set of parameters).

Analysis

All trial designs were assumed to be group sequential up to four interim analyses before a final analysis. Time-to-infection and time-to-symptoms were the outcomes of interest for ZIKV infection and Zika Virus Disease scenarios, respectively. We analyzed these trials with Cox proportional hazards models. We analyzed simulated CZS trial data using a logistic regression model with CZS diagnosis present at birth as the endpoint. We calculated statistical power as the proportion of scenario iterations in which the vaccine efficacy p-value was less than the significance level of 0.05. Mean trial duration was calculated for all trials with power of 80% or more.

Results

CZS-based clinical endpoints

For trials of vaccine efficacy against CZS outcomes, the relationship between size of the recruited population and power varied with the timing of the trial in the course of the epidemic (Table 4.2, Fig. 4.2). Of the three starting rules, those trials following rule 1 began when a country reached a weekly rate of 25 cases per 100,000 persons and achieved 80% statistical power at the lowest sample size. Interestingly, trials begun a week after the peak of an epidemic (rule 3) required an intermediate sample size, while those trials beginning when a country reached a higher weekly infection rate of 50 cases per 100,000 persons (rule 2) required the largest sample sizes to reach 80% power (Table 4.2, Fig.4. 2).

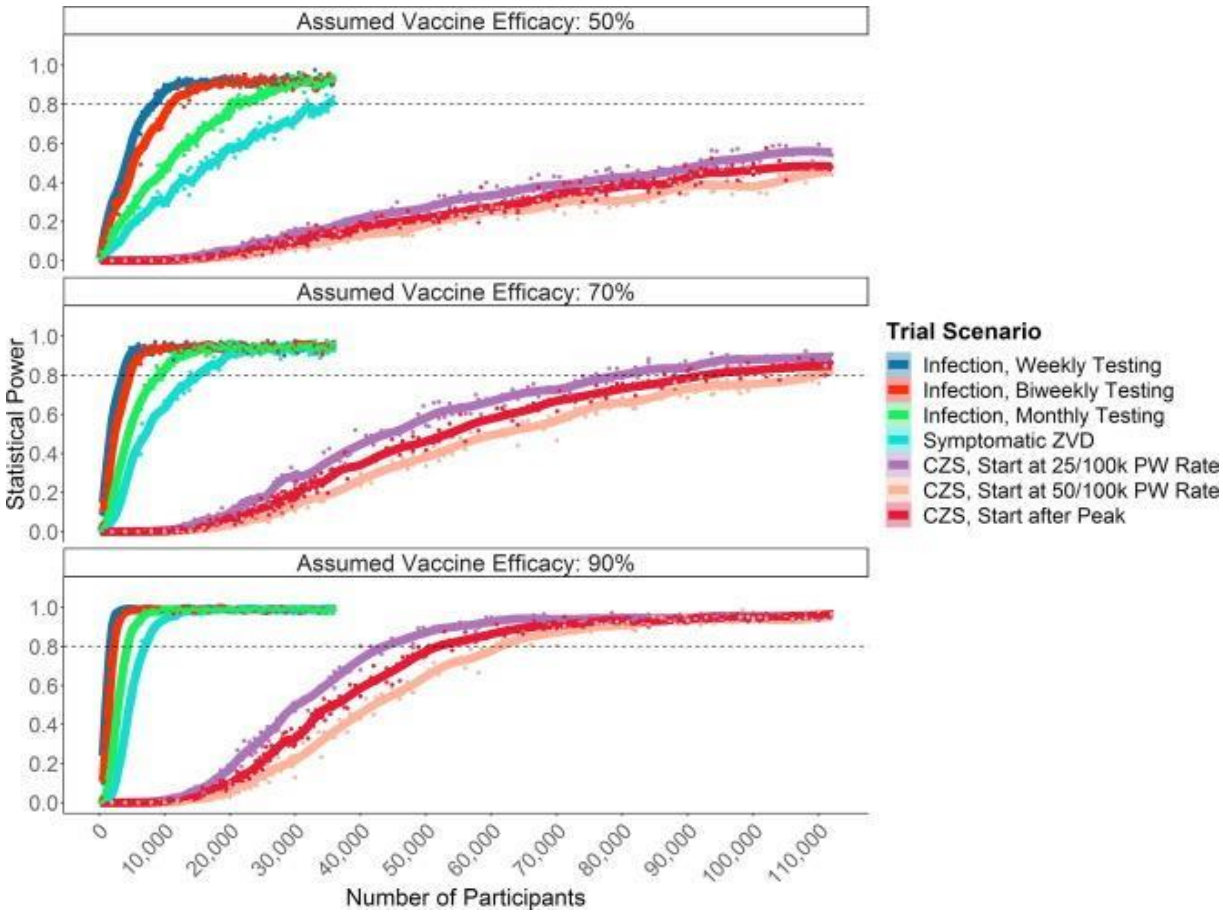

Figure 4.2**:** Statistical power to detect vaccine efficacy by number of recruited participants for seven different clinical trial scenarios and three levels of assumed vaccine efficacy.

For all starting rules, increasing assumed vaccine efficacy from 70% to 90% resulted in lower participant numbers. No trial in the 50% assumed efficacy scenarios achieved adequate power with sample sizes of up to 120,000 participants.

To better understand the effect of starting conditions on sample size for trials with CZS-based endpoints, we plotted the number of infections that occurred before completion of the month-long ramp-up period, after which the vaccine was considered protective (Fig. 4.3, top). These infections are removed in the simulation to accurately compare the vaccine and control groups. High infection rates in the first month of the trial appear to lead to more infections being excluded from the remainder of the trial analysis. Subsequently, as women are more likely to

conceive at the beginning of the trial, the number of first-trimester infections is also lower (Fig. 4.3, bottom). Trial simulations following starting rule 2 appear to have more infections removed during the pre-immunity period, and so also have fewer first-trimester infections recorded.

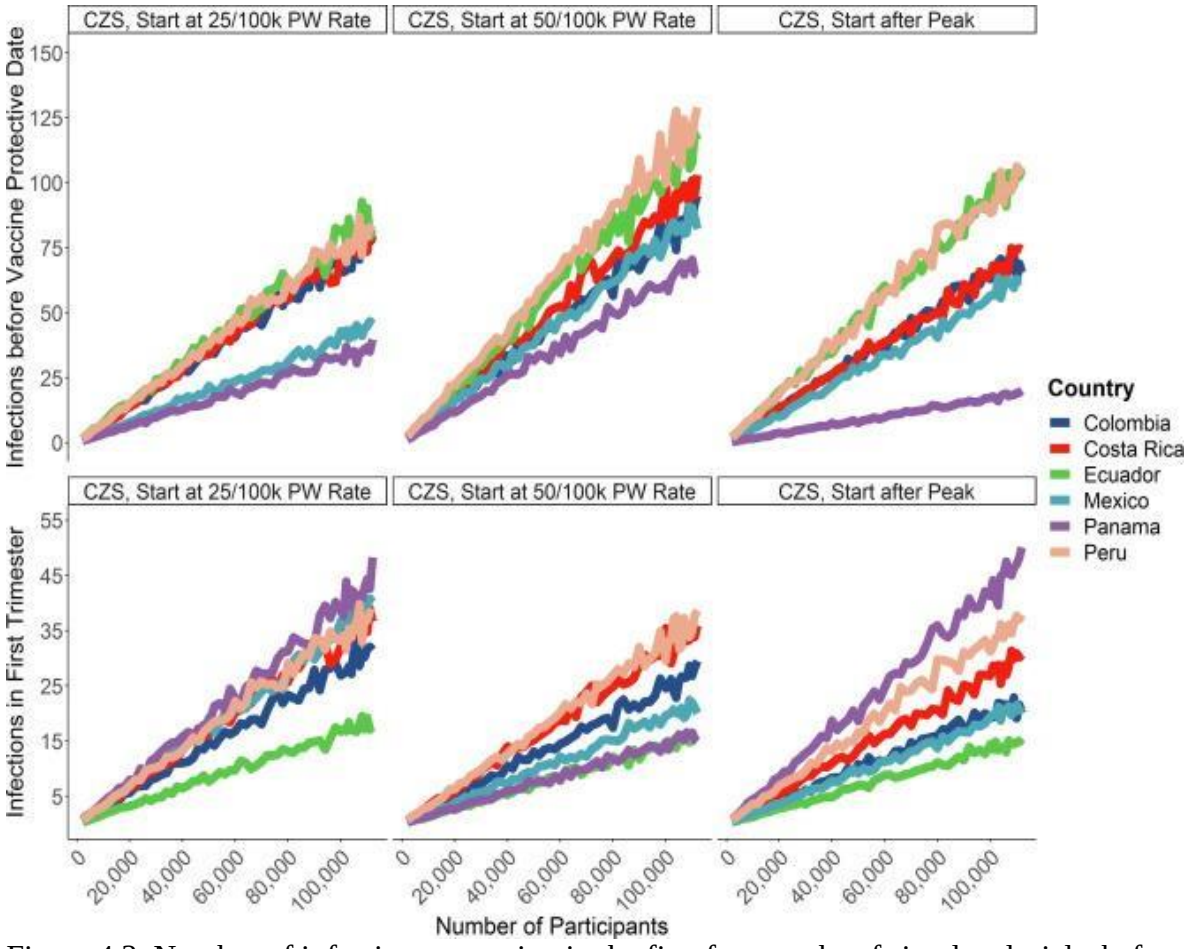

Figure 4.3. Number of infections occurring in the first four weeks of simulated trials, before the vaccine was assumed to become protective, and number of infections during the first trimester of pregnancy.

ZIKV infection- and symptomatic disease-based endpoints

For trials of vaccine efficacy against ZIKV infection or symptomatic Zika Virus Disease, the timing of the trial in the course of the epidemic did not alter sample size requirements, and so data for only one starting rule was analyzed further. In all cases, trial scenarios achieved 80%

power for far fewer participants than the number required for the CZS endpoint (Table 4.2, Fig. 4.2). For ZIKV infection, we also studied the effect of lengthening the time between routine testing visits, from weekly to biweekly or monthly. As expected, trial scenarios with a ZIKV infection endpoint identified through weekly testing were adequately powered with the fewest participants in all simulation scenarios. Doubling the time between routine tests to biweekly increased the required sample size by 26%, averaged over assumed vaccine efficacies. Approximately 2.4-times as many participants as in trials with weekly testing were needed when we increased the testing interval to monthly. For symptomatic Zika Virus Disease, sample sizes were 64% greater, on average, than those required for infection-based endpoints with monthly testing.

The length of trials, in terms of shortest mean study duration in simulations, changed little between infection and disease outcomes (Table 2). Assuming that trials recruit the minimum sample size found in simulations, extending the time between testing or moving to the symptomatic Zika Virus Disease endpoint did not appear to dramatically increase the length of the overall study period. For example, within trial scenarios with assumed vaccine efficacy of 70%, moving from weekly testing to biweekly testing increased the minimum sample size by 907 participants, but only increased the mean study period by twenty days. In trials with monthly testing, the mean duration was six days less than in trials with biweekly testing.

Discussion

CZS is the most severe outcome of ZIKV infection. As an endpoint for a ZIKV vaccine trial, it ensures that vaccine efficacy is shown against the most serious outcome of infection. While ZIKV incidence rates are low, however, CZS endpoints may not be a feasible option for vaccine efficacy trials.

Perhaps surprisingly, simulated trials with CZS endpoints in our study did achieve adequate statistical power to detect vaccine efficacy in some scenarios. These particular trials were those simulated with a start date earlier in the epidemic or a week after the peak of an epidemic. Sample sizes were greater at the intermediate start time, when incidence was highest. This result is tied to the nature of CZS risk, which varies with the trimester of infection. Probability of conception is greatest in the first months of regular sexual activity, and it was assumed women were within the first six months of conception attempts at the start of the trial. This indicates that the majority of trial pregnancies will begin near the trial start date. When this is also a period of high ZIKV transmission, infection in the first trimester and, subsequently, CZS outcomes are more likely. However, we assumed the vaccine would not yet be protective in the first four weeks of a trial, and all infections occurring in those weeks were removed from the final dataset. Figure 4.3 illustrates the consequences of these dynamics in our simulation. Those trials that began at the intermediate start time, when a country reached a 50/100,000 person-weeks infection rate, have a greater number of infections removed during the pre-immunity period, and fewer first trimester infections recorded.

Comparing the feasibility of endpoints is thus further complicated by the nature of CZS risk. As with other TORCH pathogens, the risk of adverse outcomes varies with trimester of maternal infection. For example, maternal cytomegalovirus infection is less likely to be transmitted to a fetus in early pregnancy but, given fetal infection, more likely to lead to disabilities or other symptoms at birth [93]. The complex relationships between trimester of infection and overall risk must be considered when selecting congenital infection outcomes as trial endpoints.

All CZS-endpoint simulations relied on several important assumptions, such as the ability to identify women trying to conceive, the point process with all individuals recruited and, if applicable, vaccinated at the start of the trial—thereby ensuring all women entered the trial at a period of similar ZIKV incidence—and the determination of CZS outcomes at birth. Delayed manifestations of other congenital syndromes, namely congenital rubella syndrome, are well documented [94]. Mounting evidence suggests CZS may also manifest some time after birth [95,96]. It would be necessary to extend follow-up time in vaccine trials to avoid missing delayed CZS cases.

We focused simulations on populations of women who were not already pregnant at the time of recruitment. We assumed this because of the potential delay in recognizing pregnancy, as well as the month we assumed for immune ramp-up before the vaccine was protective. These factors prevent inclusion of the entire first trimester, when CZS risk is greatest. We recognize that recruiting women who are planning on becoming pregnant presents serious ethical concerns. Choosing this population for our simulation study, however, demonstrates the complex nature of congenital infection that should be carefully considered when designing trials of vaccine efficacy against teratogenic pathogens.

We worked with eighty-two weeks of seasonal risks and assumed all pregnancies and infections must occur by the final week, regardless of how we varied the starting time of the trial. As a result, some trial scenarios ran for the entirety of the eighty-two weeks, while others ran for a shorter period. It can be expected that sample sizes would be greater for trials of shorter duration, as we artificially ended ZIKV transmission at the final week despite of any pregnancies that were not completed by that time. This eliminated any chance of infection for some pregnancies, as well as chance of multiple infections for others. Fewer pregnancies may be

expected in later trial weeks, however, as we assumed all women entered the trial at an average probability of conception, and this probability declined significantly by the end of the simulated trials.

Despite these limitations, simulation results revealed a critical relationship between ZIKV incidence at the beginning of a trial and the feasibility of a CZS clinical endpoint. In the event a trial with CZS endpoints remains infeasible regardless, infection- and disease-based endpoints are alternatives. Trial duration varied little between the infection endpoint scenarios, regardless of time between laboratory tests (Table 4.2). Trials that extend the time between sampling, though recruiting more individuals, halve or quarter the laboratory tests required *per participant*. This decreases participant inconvenience and potentially improves protocol adherence. These endpoints also avoid the need for intense trial surveillance to identify mild disease cases, which is recommended for trials that rely on symptomatic Zika Virus Disease endpoints [82]. Disease endpoints also required an additional 64% increase in participant numbers over monthly-testing ZIKV infection trial scenarios.

In terms of laboratory testing, we did not vary the sensitivity of tests to compare urine versus blood testing methods. Case identification was assumed to rely only on performing the test within the period of detectable RNA. We assumed that trials will follow current recommended laboratory protocols, which for ZIKV diagnosis currently requires paired blood and urine testing [82,85,86]. As the difference in the RNA detection interval between these sample types remains debatable, we used blood testing parameters to simplify the simulation. These tests have lower limits of detection that may be higher than the viral load required for maternal-fetal ZIKV transmission, so we recommend caution in interpreting maternal infection-based trial results in terms of CZS risk.

In all simulations, we assumed the vaccine would be protective after one month. However, this period was kept constant in all simulations and did not vary between endpoints. Similarly, as we relied on an assumed annual cumulative incidence rate, the qualitative patterns between trial scenarios will be more robust than sample sizes or trial durations provided by the simulation (Table 4.2). Actual values will depend on the vaccine under study and on incidence during the trial period.

In conclusion, though laboratory-confirmed symptomatic Zika Virus Disease and infection-based endpoints are perhaps more feasible, we found CZS clinical trial endpoints were possible in several simulated trials due to the relationship between CZS risk at trimester of infection and the timing of the trial in the course of an epidemic. This result highlights additional considerations needed when designing studies with pregnancy-related clinical endpoints, including studies involving other teratogenic pathogens.

Table 4.1. Simulation parameters.

Parameter	**Value**	**Ref.**
Incubation period in days, Median (25%, 75%)	5.9 (4.6, 7.6)	(85)
Persistence of viremia in days, Mean (25%, 75%)	9.9 (5.8, 12.7)	(85)
CZS risk, 1st trimester infection	0.15	(88)
CZS risk, 2nd/3rd trimester infection	0.0227	(88)
CZS risk, infection during multiple trimesters, including 1st	0.21	(88)
Contraceptive use, proportion of Latin American women aged 15–49 years	73%	(89)
Recruited population size, assumed vaccine efficacy, conception probability, start date within scaled epidemic data	Varied	

Table 4.2. Minimum required sample size and study duration for trial scenarios with 80% power.

	Assumed efficacy: 50%		Assumed efficacy: 70%		Assumed efficacy: 90%	
Trial scenario endpoint	**N**	**Duration (days)**	**N**	**Duration (days)**	**N**	**Duration (days)**
Infection based						
Weekly testing	8357	296	3825	287	2013	294
Biweekly testing	11,076	306	4732	307	2466	313
Monthly testing	21,046	307	9263	301	4732	321
Symptom-based						
Symptomatic ZVD	35,094	316	15,608	310	7451	324
CZS-based						
Start at 25/100k pw rate	–	–	79,451	598	44,071	626
Start at 50/100k pw rate	–	–	110,585	507	62,468	530
Start after peak	–	–	93,603	515	52,562	546

CHAPTER 5

CONCLUSION

In the previous chapters of this dissertation, results from three independent studies have illustrated how the epidemiology of an infectious disease is affected by dynamics at numerous biological scales of organization.

After SARS-CoV-2 was identified as the causal organism responsible for the outbreak of severe, pneumonia-like, disease in China, researchers around the world made an unprecedented effort to better understand its transmission and disease severity. Meanwhile, economists asked about economic impacts of stay-at-home orders [97,98], and psychologists stressed the importance of human contact in mental health [99]. While seemingly separate issues, each of these facets of the overall COVID-19 response loop back to basic infectious disease questions through better understanding of human behavior: who is getting sick and where are they infected? Broad-scale determinants of mortality, however, are more difficult to discern when individual-level risk factors are more obvious.

In the COVID-19 study presented here, age, race, availability of medical resources, and even climate factors are associated with the risk of death from COVID-19, regardless of transmission rates. The scales at which these factors affect the pandemic most likely include the entire range of biological organization: molecular processes that degrade with age or are already modified by race interact with ecosystem-level climatological factors. In our analysis, data is aggregated by county and can only be interpreted at broader scales, the narrowest of which is population. As an example, each county is associated with a certain population density that is

highly correlated with the resources available for the care of COVID-19 patients—larger population sizes indicate more medical establishments, hospitals, and highly-trained professionals, while more rural, sparsely populated counties have no hospitals and a population that must travel to receive care. Ecosystem-level climatological factors were also found significant in this analysis, with higher temperatures and relative humidity associated with reduced mortality.

Individual Tuberculosis (TB) patients in Kampala, Uganda, have diverse outcomes dependent on their fellow population and community members. As shown in the TB disease study herein, the training of the individuals with whom an individual patient primarily interacts is significantly associated with that patient's treatment. The amount of time a patient remains untreated, while actively seeking care, increases with the proportion of his or her contacts who are primary-level health providers or local healers. Potential intervention efforts highlight the interacting nature of individuals and their community. Education efforts aimed at one population—the health providers—will have far reaching consequences for the entire community. Should these health providers begin referring patients to TB diagnostic locations with greater regularity, not only will individuals receive care sooner, but their infectious period will be shortened, reducing further transmission to his or her neighbors.

The interaction of scales at the individual, population, community, and ecosystem levels may even affect the statistical power of epidemiological studies. As illustrated in the Zika virus chapter of this dissertation, individual pregnancy probability, which decreases with the time couples have attempted pregnancy, is an important component of a clinical trial for a vaccine against an infectious teratogen. As a mosquito-borne disease, Zika risk increases when seasonal mosquito activity is high. When pregnancy occurs during this time of high risk, the probability of

infections in the first trimester is increased, further increasing the risk of Congenital Zika Syndrome, or CZS. As statistical power to detect vaccine efficacy depends on the number of cases observed, power could be expected to be reached far faster in this high-risk scenario. However, vaccines take time for full effectiveness, with any infections before the time of efficacy being deleted from analysis. This culminates in a surprising reality—it is better, in terms of resources to run a trial, to begin a trial before rates are too high. As it is not possible to exactly predict when rates will peak, it may even be beneficial to begin a trial *after* a period of high risk, when rates are beginning to decrease.

It is possible to discern a general trend from these specific studies. Not only do they illustrate the importance of incorporating various scales of biological organization to understand the etiology of disease, but that doing so impacts the practice of epidemiology and design of studies and trials. Researchers must be more aware of possible interacting scales when determining the variables they wish to collect or include in analysis, and policy should be suggested with a better understanding of how an individual affects his or her community, or vice versa. In a time of great ecosystem change due to global climate change, infectious disease epidemiologists must be even more aware of large-scale factors impacting disease emergence, while also remaining cognizant of how increased temperature may impact a single individual's immunity [100].

The scientific community has become more and more interdisciplinary in recent years, and the progress of technology has allowed for collaborations that would be impossible a mere decade ago [6]. With these resources and an interdisciplinary mindset, researchers will be able to incorporate the information of multiple scales in their studies, and be better prepared to face the threat of emerging pathogens.

REFERENCES

[1] CDC. Principles of Epidemiology | Lesson 1 - Section 10. Cdc 2020.

[2] D Coggon GR and DB. Epidemilogy for the uninitiated. Chapter 1. What is epidemiology? BMJ 1997.

[3] Marçôa R. Tuberculosis and gender – Factors influencing the risk of tuberculosis among men and women by age group. Pulmonology 2018. doi:10.1016/j.pulmoe.2018.03.004.

[4] CDC. Coronavirus Disease 2019 (COVID-19): Older Adults. Https://WwwCdcGov/Coronavirus/2019-Ncov/Need-Extra-Precautions/Older-AdultsHtml 2020.

[5] Reynolds MR, Jones AM, Petersen EE, Lee EH, Rice ME, Bingham A, et al. Vital Signs: Update on Zika Virus–Associated Birth Defects and Evaluation of All U.S. Infants with Congenital Zika Virus Exposure — U.S. Zika Pregnancy Registry, 2016. MMWR Morb Mortal Wkly Rep 2017. doi:10.15585/mmwr.mm6613e1.

[6] Finucane ML, Fox J, Saksena S, Spencer JH. A conceptual framework for analyzing social-ecological models of emerging infectious diseases. Underst. Soc. Nat. Resour. Forg. New Strands Integr. Across Soc. Sci., 2014. doi:10.1007/978-94-017-8959-2_5.

[7] Wilcox BA, Colwell RR. Emerging and reemerging infectious diseases: Biocomplexity as an interdisciplinary paradigm. Ecohealth, 2005. doi:10.1007/s10393-005-8961-3.

[8] Colwell R. Balancing the biocomplexity of the planet's living systems: A twenty-first century task for science. Bioscience 1998. doi:10.2307/1313389.

[9] Colwell RR. Global climate and infectious disease: The cholera paradigm. Science (80-.

)., 1996. doi:10.1126/science.274.5295.2025.

[10] Hirsch EB, Tam VH. Impact of multidrug-resistant Pseudomonas aeruginosa infection on patient outcomes. Expert Rev Pharmacoeconomics Outcomes Res 2010. doi:10.1586/erp.10.49.

[11] Romling U, Wingender J, Muller H, Tummler B. A major Pseudomonas aeruginosa clone common to patients and aquatic habitats. Appl Environ Microbiol 1994. doi:10.1128/aem.60.6.1734-1738.1994.

[12] Köck R, Daniels-Haardt I, Becker K, Mellmann A, Friedrich AW, Mevius D, et al. Carbapenem-resistant Enterobacteriaceae in wildlife, food-producing, and companion animals: a systematic review. Clin Microbiol Infect 2018. doi:10.1016/j.cmi.2018.04.004.

[13] Scott A, Pottenger S, Timofte D, Moore M, Wright L, Kukavica-Ibrulj I, et al. Reservoirs of resistance: Polymyxin resistance in veterinary-associated companion animal isolates of Pseudomonas aeruginosa. Vet Rec 2019. doi:10.1136/vr.105075.

[14] Ramey AM, Ahlstrom CA. Antibiotic resistant bacteria in wildlife: Perspectives on trends, acquisition and dissemination, data gaps, and future directions. J Wildl Dis 2020. doi:10.7589/2019-04-099.

[15] Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020. doi:10.1016/S1473-3099(20)30120-1.

[16] Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China. JAMA 2020. doi:10.1001/jama.2020.2648.

[17] Franch-Pardo I, Napoletano BM, Rosete-Verges F, Billa L. Spatial analysis and GIS in the study of COVID-19. A review. Sci Total Environ 2020. doi:10.1016/j.scitotenv.2020.140033.

[18] Sugg MM, Spaulding TJ, Lane SJ, Runkle JD, Harden SR, Hege A, et al. Mapping community-level determinants of COVID-19 transmission in nursing homes: A multi-scale approach. Sci Total Environ 2021. doi:10.1016/j.scitotenv.2020.141946.

[19] Andersen LM, Harden SR, Sugg MM, Runkle JD, Lundquist TE. Analyzing the spatial determinants of local Covid-19 transmission in the United States. Sci Total Environ 2021. doi:10.1016/j.scitotenv.2020.142396.

[20] Desjardins MR, Hohl A, Delmelle EM. Rapid surveillance of COVID-19 in the United States using a prospective space-time scan statistic: Detecting and evaluating emerging clusters. Appl Geogr 2020. doi:10.1016/j.apgeog.2020.102202.

[21] Allcott H, Boxell L, Conway J, Gentzkow M, Thaler M, Yang D. Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic. J Public Econ 2020. doi:10.1016/j.jpubeco.2020.104254.

[22] Wang B, Zhang A, Sun JL, Liu H, Hu J, Xu LX. Study of SARS transmission via liquid droplets in air. J Biomech Eng 2005. doi:10.1115/1.1835350.

[23] Yuan J, Yun H, Lan W, Wang W, Sullivan SG, Jia S, et al. A climatologic investigation of the SARS-CoV outbreak in Beijing, China. Am J Infect Control 2006. doi:10.1016/j.ajic.2005.12.006.

[24] Baker RE, Yang W, Vecchi GA, Metcalf CJE, Grenfell BT. Susceptible supply limits the role of climate in the early SARS-CoV-2 pandemic. Science (80-) 2020. doi:10.1126/science.abc2535.

[25] Dublineau A, Batéjat C, Pinon A, Burguière AM, Leclercq I, Manuguerra JC. Persistence of the 2009 pandemic influenza a (H1N1) virus in water and on non-porous surface. PLoS One 2011. doi:10.1371/journal.pone.0028043.

[26] Shaman J, Pitzer VE, Viboud C, Grenfell BT, Lipsitch M. Absolute humidity and the seasonal onset of influenza in the continental United States. PLoS Biol 2010. doi:10.1371/journal.pbio.1000316.

[27] Shaman J, Kohn M. Absolute humidity modulates influenza survival, transmission, and seasonality. Proc Natl Acad Sci U S A 2009. doi:10.1073/pnas.0806852106.

[28] Paynter S. Humidity and respiratory virus transmission in tropical and temperate settings. Epidemiol Infect 2015. doi:10.1017/S0950268814002702.

[29] Kudo E, Song E, Yockey LJ, Rakib T, Wong PW, Homer RJ, et al. Low ambient humidity impairs barrier function and innate resistance against influenza infection. Proc Natl Acad Sci U S A 2019. doi:10.1073/pnas.1902840116.

[30] Kifer D, Bugada D, Villar-Garcia J, Gudelj I, Menni C, Sudre C, et al. Effects of Environmental Factors on Severity and Mortality of COVID-19. Front Med 2021. doi:10.3389/fmed.2020.607786.

[31] Courtney JM, Bax A. Hydrating the respiratory tract: An alternative explanation why masks lower severity of COVID-19 disease. MedRxiv 2020. doi:10.1101/2020.12.23.20248671.

[32] The New York Times. Coronavirus (Covid-19) Data in the United States. Github 2020.

[33] R Core Team. R: A language and environment for statistical computing. R Found Stat Comput 2019.

[34] H Sparks A, Hengl T, Nelson A. GSODR: Global Summary Daily Weather Data in R. J Open Source Softw 2017. doi:10.21105/joss.00177.

[35] Olmedo OE. Kriging: Ordinary Kriging 2014.

[36] USCB. US Census Bureau 2010 Census. 2016 2016.

[37] Gundersen C, Dewey A, Engelhard E, Strayer M, Lapinski L. Map the Meal Gap 2020: A Report on County and Congressional District Food Insecurity and County Food Cost in the United States in 2018 2020.

[38] ESRI. ArcGIS Desktop. ArcGIS Blog 2018.

[39] SaTScan. SaTScan - Software for the spatial, temporal, and space-time scan statistics. Https://WwwSatscanOrg/ 2005.

[40] Malizia N. Inaccuracy, Uncertainty and the Space-Time Permutation Scan Statistic. PLoS One 2013. doi:10.1371/journal.pone.0052034.

[41] World Health Organisation. Global Health TB Report. 2018. doi:ISBN 978-92-4-156564-6.

[42] Onozaki I, Law I, Sismanidis C, Zignol M, Glaziou P, Floyd K. National tuberculosis prevalence surveys in Asia, 1990-2012: An overview of results and lessons learned. Trop Med Int Heal 2015. doi:10.1111/tmi.12534.

[43] Storla DG, Yimer S, Bjune GA. A systematic review of delay in the diagnosis and treatment of tuberculosis. BMC Public Health 2008. doi:10.1186/1471-2458-8-15.

[44] Madebo T, Lindtjorn B. Delay in Treatment of Pulmonary Tuberculosis: An Analysis of Symptom Duration Among Ethiopian Patients. MedGenMed 1999.

[45] Sekandi JN, Zalwango S, Martinez L, Handel A, Kakaire R, Nkwata AK, et al. Four Degrees of Separation: Social Contacts and Health Providers Influence the Steps to Final Diagnosis of Active Tuberculosis Patients in Urban Uganda. BMC Infect Dis 2015. doi:10.1186/s12879-015-1084-8.

[46] Handel A, Martinez L, Sekandi JN, Bellan SE, Zhu L, Chen C, et al. Evidence for supercoughers in an analysis of six tuberculosis cohorts from China, Peru, the Gambia and

Uganda. Int J Tuberc Lung Dis 2019. doi:10.5588/ijtld.18.0819.

[47] Saifodine A, Gudo PS, Sidat M, Black J. Patient and health system delay among patients with pulmonary tuberculosis in Beira city, Mozambique. BMC Public Health 2013. doi:10.1186/1471-2458-13-559.

[48] Rabin AS, Kuchukhidze G, Sanikidze E, Kempker RR, Blumberg HM. Prescribed and self-medication use increase delays in diagnosis of tuberculosis in the country of Georgia. Int J Tuberc Lung Dis 2013. doi:10.5588/ijtld.12.0395.

[49] Ukwaja KN, Alobu I, Nweke CO, Onyenwe EC. Healthcare-seeking behavior, treatment delays and its determinants among pulmonary tuberculosis patients in rural Nigeria: a cross-sectional study. BMC Health Serv Res 2013. doi:10.1186/1472-6963-13-25.

[50] Huong NT, Vree M, Duong BD, Khanh VT, Loan VT, Co N V., et al. Delays in the diagnosis and treatment of tuberculosis patients in Vietnam: A cross-sectional study. BMC Public Health 2007. doi:10.1186/1471-2458-7-110.

[51] Belkina T V., Khojiev DS, Tillyashaykhov MN, Tigay ZN, Kudenov MU, Tebbens JD, et al. Delay in the diagnosis and treatment of pulmonary tuberculosis in Uzbekistan: A cross-sectional study. BMC Infect Dis 2014. doi:10.1186/s12879-014-0624-y.

[52] Takarinda KC, Harries AD, Nyathi B, Ngwenya M, Mutasa-Apollo T, Sandy C. Tuberculosis treatment delays and associated factors within the Zimbabwe national tuberculosis programme. BMC Public Health 2015. doi:10.1186/s12889-015-1437-7.

[53] Ayé R, Wyss K, Abdualimova H, Saidaliev S. Patient's site of first access to health system influences length of delay for tuberculosis treatment in Tajikistan. BMC Health Serv Res 2010. doi:10.1186/1472-6963-10-10.

[54] Makwakwa L, Sheu M ling, Chiang CY, Lin SL, Chang PW. Patient and heath system

delays in the diagnosis and treatment of new and retreatment pulmonary tuberculosis cases in Malawi. BMC Infect Dis 2014. doi:10.1186/1471-2334-14-132.

[55] Nasehi M, Hassanzadeh J, Rezaianzadeh A, Zeigami B, Tabatabaee H, Ghaderi E. Diagnosis delay in smear positive tuberculosis patients. J Res Med Sci 2012.

[56] Gosoniu GD, Ganapathy S, Kemp J, Auer C, Somma D, Karim F, et al. Gender and socio-cultural determinants of delay to diagnosis of TB in Bangladesh, India and Malawi. Int J Tuberc Lung Dis 2008.

[57] Ford CM, Bayer AM, Gilman RH, Onifade D, Acosta C, Cabrera L, et al. Factors associated with delayed tuberculosis test-seeking behavior in the Peruvian Amazon. Am J Trop Med Hyg 2009. doi:10.4269/ajtmh.2009.08-0627.

[58] Pradhan A, Kielmann K, Gupte H, Bamne A, Porter JDH, Rangan S. What “outliers” tell us about missed opportunities for tuberculosis control: A cross-sectional study of patients in Mumbai, India. BMC Public Health 2010. doi:10.1186/1471-2458-10-263.

[59] Getnet F, Demissie M, Assefa N, Mengistie B, Worku A. Delay in diagnosis of pulmonary tuberculosis in low-and middle-income settings: Systematic review and meta-analysis. BMC Pulm Med 2017. doi:10.1186/s12890-017-0551-y.

[60] Jurcev-Savicevic A, Mulic R, Kozul K, Ban B, Valic J, Bacun-Ivcek L, et al. Health system delay in pulmonary tuberculosis treatment in a country with an intermediate burden of tuberculosis: A cross-sectional study. BMC Public Health 2013. doi:10.1186/1471-2458-13-250.

[61] Kansiime C, Kiwuwa SM, Levi M, Asiimwe BB, Katamba A. Health service delay among pulmonary tuberculosis patients presenting to a National Referral Hospital, Kampala, Uganda: A cross sectional study. Pan Afr Med J 2013.

doi:10.11604/pamj.2013.15.84.2692.

[62] Maciel ELN, Golub JE, Peres RL, Hadad DJ, Fávero JL, Molino LP, et al. Delay in diagnosis of pulmonary tuberculosis at a primary health clinic in Vitoria, Brazil. Int J Tuberc Lung Dis 2010.

[63] Tattevin P, Che D, Fraisse P, Gatey C, Guichard C, Antoine D, et al. Factors associated with patient and health care system delay in the diagnosis of tuberculosis in France. Int J Tuberc Lung Dis 2012. doi:10.5588/ijtld.11.0420.

[64] Abebe G, Deribew A, Apers L, Woldemichael K, Shiffa J, Tesfaye M, et al. Knowledge, health seeking behavior and perceived stigma towards tuberculosis among tuberculosis suspects in a rural community in Southwest Ethiopia. PLoS One 2010. doi:10.1371/journal.pone.0013339.

[65] Tupasi TE, Radhakrishna S, Co VM, Villa MLA, Quelapio MID, Mangubat N V., et al. Bacillary disease and health seeking behavior among Filipinos with symptoms of tuberculosis: Implications for control. Int J Tuberc Lung Dis 2000.

[66] Sreeramareddy CT, Qin ZZ, Satyanarayana S, Subbaraman R, Pai M. Delays in diagnosis and treatment of pulmonary tuberculosis in India: A systematic review. Int J Tuberc Lung Dis 2014. doi:10.5588/ijtld.13.0585.

[67] USAID. Health facility census in the Kampala capital city authority (KCCA) divisions - census report. 2017.

[68] Copas AJ. Incorporating retrospective data into an analysis of time to illness. Biostatistics 2001. doi:10.1093/biostatistics/2.1.1.

[69] Sendagire I, Van der Loeff MS, Mubiru M, Konde-Lule J, Cobelens F. Long delays and missed opportunities in diagnosing smear-positive pulmonary tuberculosis in Kampala,

Uganda: A cross-sectional study. PLoS One 2010. doi:10.1371/journal.pone.0014459.

[70] Yadav S, Rawal G. The novel concept of creating awareness about tuberculosis at the metro stations. Pan Afr Med J 2016. doi:10.11604/pamj.2016.23.228.7127.

[71] García-Basteiro AL, DiNardo A, Saavedra B, Silva DR, Palmero D, Gegia M, et al. Point of care diagnostics for tuberculosis. Rev Port Pneumol (English Ed 2018. doi:10.1016/j.rppnen.2017.12.002.

[72] Petersen LR, Jamieson DJ, Powers AM, Honein MA. Zika Virus. N Engl J Med 2016;374:1552–63. doi:10.1056/NEJMra1602113.

[73] Rasmussen SA, Jamieson DJ, Honein MA, Petersen LR. Zika Virus and Birth Defects — Reviewing the Evidence for Causality. N Engl J Med 2016;374:1981–7. doi:10.1056/NEJMsr1604338.

[74] Cao-Lormeau VM, Blake A, Mons S, Lastere S, Roche C, Vanhomwegen J, et al. Guillain-Barre Syndrome outbreak associated with Zika virus infection in French Polynesia: a case-control study. Lancet 2016;387:1531–9. doi:10.1016/S0140-6736(16)00562-6.

[75] European Centre for Disease Prevention and Control. Rapid risk assessment: Zika virus infection outbreak, French Polynesia. ECDC 2014:1–12.

[76] Mehrjardi MZ. Is Zika virus an emerging TORCH agent? An invited commentary. Virol Res Treat 2017. doi:10.1177/1178122X17708993.

[77] Pan American Health Organization. Countries and territories with autochthonous transmission in the Americas reported in 2015-2017 2018:41696. https://www.paho.org/hq/index.php?option=com_content&view=article&id=11603&Itemid=41696&lang=en.

[78] World Health Organization. WHO Director-General summarizes the outcome of the Emergency Committee regarding clusters of microcephaly and Guillain-Barré syndrome - WHO Statement 2016.

[79] World Health Organization. Fifth meeting of the Emergency Committee under the International Health Regulations (2005) regarding microcephaly, other neurological disorders and Zika virus - WHO Statement 2016.

[80] Pardi N, Hogan MJ, Pelc RS, Muramatsu H, Andersen H, DeMaso CR, et al. Zika virus protection by a single low-dose nucleoside-modified mRNA vaccination. Nature 2017;543:248–51. doi:10.1038/nature21428.

[81] Asher J, Barker C, Chen G, Cummings D, Chinazzi M, Daniel-Wayman S, et al. Preliminary modeling results for Zika virus transmission in 2017. BioRxiv 2017:1–26. doi:http://dx.doi.org/10.1101/187591.

[82] World Health Organization. "Efficacy trials of ZIKV Vaccines: endpoints, trial design, site selection" - WHO Workshop Meeting Report. Geneva: 2017.

[83] Zorrilla CD, García García I, García Fragoso L, De La Vega A. Zika Virus Infection in Pregnancy: Maternal, Fetal, and Neonatal Considerations. J Infect Dis 2017. doi:10.1093/infdis/jix448.

[84] Barouch DH, Thomas SJ, Michael NL. Prospects for a Zika Virus Vaccine. Immunity 2017;46:176–82. doi:10.1016/j.immuni.2017.02.005.

[85] Centers for Disease Control. Testing Guidance. CdcGov 2018. https://www.cdc.gov/zika/hc-providers/testing-guidance.html.

[86] Centers for Disease Control. Guidance for US Laboratories Testing for Zika Virus Infection 2017:1–16. https://www.cdc.gov/zika/pdfs/laboratory-guidance-zika.pdf.

[87] Theel ES, Jane Hata D. Diagnostic testing for Zika virus: A postoutbreak update. J Clin Microbiol 2018. doi:10.1128/JCM.01972-17.

[88] Andersen K. Zika case numbers from PAHO 2018. https://andersen-lab.com/secrets/data/zika-paho-numbers/.

[89] Lessler J, Ott CT, Carcelen AC, Konikoff JM, Williamson J, Bi Q, et al. Times to Key Events in the Course of Zika Infection and their Implications for Surveillance: A Systematic Review and Pooled Analysis. Igarss 2014 2016. doi:10.1101/041913.

[90] Justus CG, Hargraves WR, Mikhail A, Graber D. Methods for Estimating Wind Speed Frequency Distributions. J Appl Meteorol 1978. doi:10.1175/1520-0450(1978)017<0350:MFEWSF>2.0.CO;2.

[91] Taylor A. Extent of the problem. BMJ 2003;327:434. doi:10.1136/bmj.327.7412.434.

[92] United Nation, Department of Economic and Social Affairs PD. Trends in contraceptive use Worldwide 2015. 2015. doi:10.1016/j.contraception.2012.08.029.

[93] Pass RF, Anderson B. Mother-to-child transmission of cytomegalovirus and prevention of congenital infection. J Pediatric Infect Dis Soc 2014. doi:10.1093/jpids/piu069.

[94] Sever JL, South MA, Shaver KA. Delayed manifestations of congenital rubella. Rev Infect Dis 1985.

[95] Eppes C, Rac M, Dunn J, Versalovic J, Murray KO, Suter MA, et al. Testing for Zika virus infection in pregnancy: key concepts to deal with an emerging epidemic. Am J Obstet Gynecol 2017. doi:10.1016/j.ajog.2017.01.020.

[96] van der Linden V, Pessoa A, Dobyns W, Barkovich AJ, Júnior H van der L, Filho ELR, et al. Description of 13 Infants Born During October 2015–January 2016 With Congenital Zika Virus Infection Without Microcephaly at Birth — Brazil. MMWR Morb Mortal

Wkly Rep 2016. doi:10.15585/mmwr.mm6547e2.

[97] Bairoliya N, Imrohoroglu A. Macroeconomic Consequences of Stay-At-Home Policies During the COVID-19 Pandemic. 2020.

[98] Baek C, McCrory PB, Messer T, Mui P. Unemployment Effects of Stay-at-Home Orders: Evidence from High Frequency Claims Data. Rev Econ Stat 2020. doi:10.1162/rest_a_00996.

[99] Marroquín B, Vine V, Morgan R. Mental health during the COVID-19 pandemic: Effects of stay-at-home policies, social distancing behavior, and social resources. Psychiatry Res 2020. doi:10.1016/j.psychres.2020.113419.

[100] Xu R, Li S, Guo S, Zhao Q, Abramson MJ, Li S, et al. Environmental temperature and human epigenetic modifications: A systematic review. Environ Pollut 2020. doi:10.1016/j.envpol.2019.113840.