MOVING PEOPLE, MOVING PATHOGENS: POPULATION MOVEMENT AND DISEASE EMERGENCE IN THE DEMOCRATIC REPUBLIC OF THE CONGO

Corinna Keeler

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Geography

Chapel Hill
2020

Approved by:

Michael Emch

Steven Meshnick

Jonathan Parr

Paul Delamater

Conghe Song

ABSTRACT

Corinna Keeler: Moving People, Moving Pathogens: Population movement and disease emergence in the Democratic Republic of the Congo
(Under the direction of Michael Emch)

This project investigates relationships between geography, demography, and infectious disease epidemiology in the Democratic Republic of the Congo (DRC). The DRC experiences frequent outbreaks of preventable and treatable diseases, and has been shown to be a corridor of infectious disease transmission between different regions of Africa. In the absence of a recent national census, there is a need for empirical evidence about demographic characteristics of the DRC, such as the population size and human movement patterns, to inform epidemiological estimates and infectious disease surveillance and control efforts.

Researchers have increasingly turned to gridded population datasets to calculate population size for the DRC; however, little work has been done evaluating concordance between population estimates obtained from different gridded datasets. I find that even though population estimates at the national level are similar between gridded population data sources, the subnational structure of the population at the province and health zone level is significantly different between LandScan and WorldPop data, and use the case study of a bednet distribution campaign to illustrate the relevance of these findings for implementing disease control measures.

Next, I characterize patterns in two distinct types of human movement in the DRC: circulatory mobility and resettlement migration. I draw on two large cross-sectional population-based surveys to construct estimates of the rates of each of these types of migration in the DRC,

and model the relationship between demographic and employment characteristics and circulatory mobility. I also examine geographic patterns in resettlement migration as well as the dominant motivations for migration among respondents in the survey dataset, given the significance of understanding population movements for measuring infectious disease outcomes.

Finally, I examine the geography and epidemiology of yellow fever virus (YFV) occurrence using data from the Integrated Disease Surveillance and Response system in the DRC. I describe geographic, temporal, and seasonal trends in YFV cases from 2005-2017 in the DRC. I then examine relationships between YFV and the demographic factors I investigated in earlier chapters, and construct statistical models of probability of YFV occurrence using environmental correlates including temperature and precipitation.

ACKNOWLEDGEMENTS

The process of spending *mumble* years in graduate school and writing a dissertation has given me a visceral understanding of the many ways in which research is a collaborative effort – in generating and refining ideas; in maintaining both personal and scholarly perspective; and in making meaning out of maps, numbers, and words. I am so grateful to my committee, my family, and my intersecting UNC communities in Geography, IDEEL, and CPC for being part of the iterative process of this project.

Mike, thank you for all of the advice and guidance you've given me, both in general and for this project, and for your willingness to debate with me about critical geography. Steve, thank you for encouraging me to run with ideas, and using the creativity and rigor that you bring to scientific inquiry to help me make this project actually happen. Jonathan, thank you for sharing your clinical knowledge and experience in the DRC to help me make this project as applied and relevant as possible. Paul, thank you for the clarity with which you think about (and explain!) spatial and statistical methods. Conghe, thank you for representing the best of how geographers can integrate spatial science with insights from the breadth of our discipline.

Thank you to my parents, Joanne and Andy, and my sister, Rachel, for listening to me alternately talk about this project and adamantly not talk about it for the past two years. I feel so fortunate to have grown up with, and had my worldview shaped by, all three of you – your values, intelligence, and Spades acumen. And thank you to Varun, Adrien, and Libby for being not only role models of thoughtful research, but also supportive, funny, and intellectually

generous friends with excellent taste in chickpeas and dogs. Libby, the best part of writing this dissertation was doing it side-by-side with you, both literally and metaphorically. But then again that's kind of my whole thing.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AFRO	World Health Organization Regional Office for Africa
AIC	Akaike Information Criterion
CDC	United States Centers for Disease Control
DHIS-2	District Health Information Software 2
DHS	Demographic and Health Survey
DON	Disease Outbreak News
DRC	Democratic Republic of the Congo
GHS	Global Human Settlement Data
GPW	Gridded Population of the World
HIV	Human Immunodeficiency Virus
HZ	Health Zone
IDP	Internally Displaced Person
IDSR	Integrated Disease Surveillance and Response
ITN	Insecticide Treated Net
MMWR	Morbidity and Mortality Weekly Report
UN	United Nations
UNHCR	United Nations High Commission on Refugees
UNOCHA	United Nations Office for the Coordination of Humanitarian Affairs
USAID	United States Agency for International Development
WHO	World Health Organization
YFV	Yellow Fever Virus

CHAPTER 1. INTRODUCTION

The World Health Organization's online archive of Disease Outbreak News alerts begins in 1996 and continues into 2020, with weekly updates regarding every disease outbreak that the WHO is currently monitoring due to concern about significant local public health impact or international spread[1]. In the 25 years represented in this online archive, the Democratic Republic of Congo (DRC) is the only country that has had disease outbreaks every single year that rise to the level of reporting in the WHO Disease Outbreak News system. The disease alerts for the DRC span viral, bacterial, and protozoan diseases; they discuss vector-borne, water-borne, and person-to-person transmission, and they range in geographic scope from documenting nationwide cholera epidemics to two cases of polio in a single health zone. No other nation has as many disease outbreak reports, regarding as many different diseases, at as many different time points.

Throughout this archive, several themes emerge regarding 25 years of disease outbreaks in the DRC. One such theme is the challenge of public health surveillance, intervention implementation, and disease control in a country as geographically large and varied as the DRC. Another significant theme that arises in the collection of Disease Outbreak News bulletins is the role of the DRC's highly mobile population in propagating and complicating disease outbreaks, with frequent mentions of refugee movements, travel for seasonal labor, and rural migration to

---

[1] For some light reading when you finish this dissertation, these archives can be accessed at https://www.who.int/csr/don/en/ and can be sorted by disease, year, or country/area.

cities. A third theme is uncertainty in population measurement of various cities and provinces in the DRC, which adds a layer of complexity to enumerating susceptible populations and estimating vaccination needs.

This dissertation project is grounded questions about geography, demography, and infectious disease epidemiology that arise repeatedly in the archive Disease Outbreak News alerts for the DRC. The uncertainty surrounding population distribution and population movement in the DRC inherently complicates public health measures and disease surveillance efforts, even as these uncertainties stem from many of the same aspects of the DRC's physical and cultural geography that render it susceptible to frequent disease outbreaks: the large and environmentally varied land area of the country, its relative political instability and ongoing civil violence, and its lack of infrastructural development compared to neighboring nations in sub-Saharan Africa. Against that backdrop, this project undertakes three interrelated aims: first, to examine the spatial distribution of the DRC population at relevant administrative units in order to provide insight into the "denominator" during disease outbreaks; second, to characterize nationwide patterns of human movements in the context of the relationship between population movement and infectious disease exposures; and finally, to examine the geography and epidemiology of yellow fever virus in the DRC in light of the demographic insights from the first two aims.

In this introduction, I describe the DRC as a study site, and include a qualitative synthesis of the WHO archive mentioned above to contextualize the relevance of my work for health policy and planning in the country. I then review the guiding insights from three distinct literatures – health geography, demography, and infectious disease epidemiology – that frame and ground this project. Finally, I introduce the structure of the dissertation, including the primary research question and approach of each chapter.

1. Study Site: The Democratic Republic of the Congo

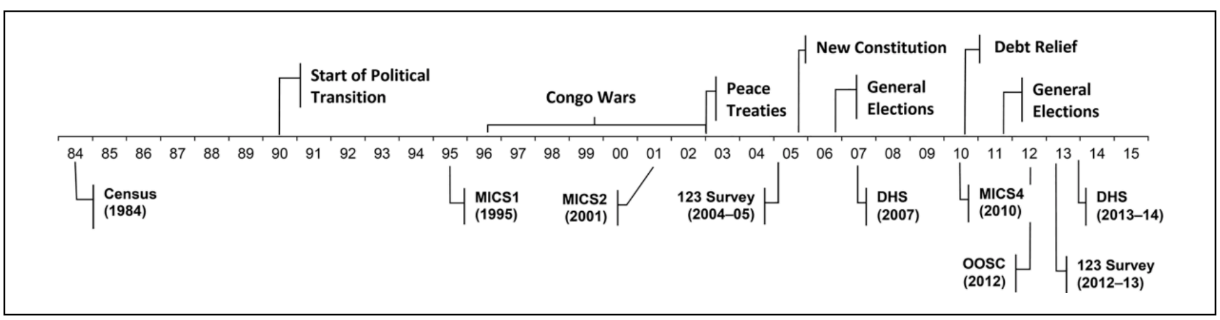
1.1 Description of study site

Perhaps it is unorthodox to start the introduction to this project with a careful examination of the study site instead of the theoretical framework or methodological approach, but there are particular characteristics of the DRC that inform every other aspect of the project and render a study site description a necessary first step. Indeed, while I can cite that the DRC is the largest country in Sub-Saharan Africa by land area, second only to Algeria on the continent (United Nations, 2016), it is difficult to make a similar comparison in terms of population because there has not been a national census since 1984. In "From Figures to Facts: Making sense of socioeconomic surveys in the DRC," Marivoet and De Herdt outline the difficulty of research, administration, and planning in the DRC without reliable population information. Their observation is so salient to the work presented in this dissertation that I am including it here in full:

> *"First of all, and most importantly, as nobody really seems to know how many Congolese today populate the DRC, any effective or efficient policy on behalf of the government or donor community is largely excluded. How many schools the Ministry of Education should build in the next five years in Kinshasa; how many vaccinations or food packages development partners should prepare for certain health and food-insecure regions; or how much budget the government should earmark for its current, future and retiring staff of civil servants, all depend on a knowledge of absolute figures. Yet, any presumed knowledge on this highly differs. Indeed, by using the population growth rates which implicitly informed the sampling frames of the latest seven national household surveys, the DRC would count today either 77.1or 93.4 million people (or some figure in-between). This difference of 16.3 million equals the estimated population of the neighboring country Zambia while 2/3rd of all countries in the world have a total population size lower than this margin of imprecision."* (Marivoet & De Herdt, 2017)

In the absence of a national census, there have been a series of cross-sectional population-based surveys in the DRC, mostly since 2000, which focus on various outcomes including out-of-school children, demographics and health, and livelihoods. The timing of these surveys is described in Figure 1. This project incorporates data from the two most recent of these national surveys, the Enquéte 1-2-3 conducted in 2012 and the Demographic and Health Survey (DHS) conducted in 2013-2014. In the years between the 1984 national census and these two surveys, the DRC experienced an ongoing period of civil and ethnic violence referred to as the Congo Wars, creating large numbers of internally displaced persons (IDPs) and refugees; violent conflict, forced displacement, involuntary movement continued to be a challenge in the country well beyond the peace treaty that putatively ended the war (Tamm, 2016; UNHCR, 2014).

Figure 1. Timeline of national cross-sectional surveys in the DRC, compared to the most recent census and key political events. Figure from Marivoet and De Herdt (2018).



The DRC shares a national border with nine other countries: Angola, Zambia, Tanzania, Burundi, Rwanda, Uganda, South Sudan, Central African Republic, and Republic of Congo. As Marivoet and De Herdt highlight in the passage above, the total population size is uncertain (a topic which is examined in detail in Chapter 2 of this dissertation), but estimates range from ~75,000,000 to ~100,000,000, making the DRC the third most populous country in Africa. The administrative divisions of the DRC are salient for this project; the country is divided into 26 provinces. The secondary political administrative unit is the territory (N=162), while the secondary health administrative unit is the health zone (N=515). The land area of the DRC is approximately 2,345,000 square kilometers, approximately 25% as large as the United States.

1.2 Disease outbreaks in the DRC

I examined the full archive of WHO Disease Outbreak News (DON) alerts in the DRC in order to understand the challenges to disease control efforts that have persisted in the DRC, and used this qualitative analysis to guide my own research. As discussed above, the DRC is the only nation to have a disease outbreak that rises to the level of reporting as a DON alert in every year of the online archive, 1996 to present. My qualitative synthesis of the major trends and themes in this archive not only positions my dissertation in relation to outstanding questions about the geography of disease outbreaks in the DRC, but also contextualizes the landscape of infectious disease research, intervention, and control throughout the country in recent history.

There are 180 DON alerts regarding the DRC in the online archive. I applied a qualitative coding scheme to each DON document, coding for the following items: the disease(s) of concern; the province, health zone, and/or geographic area affected; actors involved in disease response including government agencies/ministries, private organizations, and non-governmental organizations; whether the alert pertained to a new outbreak or was an update on a previously documented outbreak; major risk factors described in the document; major transmission pathways described in the document; and discussion of migration, mobility, or population movement.

In the time from 1996 to present, the DRC experienced outbreaks of yellow fever, cholera, meningitis, monkeypox, plague, influenza, typhoid fever, both wild type polio and vaccine-derived polio, and acute hemorrhagic fevers including Ebola and Marburg. Of the 180 DON alerts, 44 (24.4%) describe a new outbreak; the remaining 136 provided updates on those

44 disease outbreaks. The topics of the updates range from documenting further geographic expansion or increased infection rates of the outbreak, to announcing newly implemented diagnostic criteria or control measures, to broadcasting that an epidemic had concluded. In particular, I noted that there were four yellow fever virus (YFV) outbreaks documented in the DON archives; all of them were in the past 10 years, in 2010, 2013, 2014, and 2016.

The challenges introduced by the lack of census data in the DRC are reflected in the DON archive. During an outbreak of Ebola in 2018, the DONs referred to concern over the proximity of the outbreak to Mbandaka, a large city in the northern part of the DRC which is also a major port on the Congo River with travel and trade connections to both Republic of Congo and Central Africa Republic. The DON alert dated May 14, 2018 refers to the fact that the outbreak had reached "Wangata health zone [which is] adjacent to the provincial port city of Mbandaka (population 1.2 million)." However, in the following week, an update on the same outbreak dated May 17, 2018 reads: "Mbandaka City … has a population of approximately 1.5 million people." As the WHO and international partners were in the process of stockpiling vaccine supply to enact experimental ring vaccinations to control the outbreak, this discrepancy in population has clear ramifications for health planning.

A frequent theme that arises in the collection of Disease Outbreak News bulletins is the role of the DRC's highly mobile population in propagating and complicating disease outbreaks. Many of the DONs make reference to refugee movements, particularly with regard to cholera outbreaks. The mention of the relationship between refugee movements and cholera epidemics happens in nearly every year of the data, in outbreaks in at least 14 provinces as well as cross-border outbreaks. Similarly, travel for seasonal labor is mentioned in the context of both farming and mining. In the 2010 yellow fever outbreak, the index case was a farmer who returned home to a mid-sized village after several weeks away. There are eight distinct Ebola outbreaks documented in the archive, and over half identify the index case as either a farmer or a miner.

Finally, even without specific mention of the direction or cause of the mobility, the highly mobile population of the DRC is invoked in the DON archive as a reality that complicates infectious disease control in the country. A representative quote can be found in a March 21, 2019 DON update for the ongoing Ebola outbreak in the northwest part of the DRC: “Given the geographical spread of the epidemic and the high mobility in this region, the risk of Ebola spreading to unaffected areas or being reintroduced to previously affected areas remains high.” In the qualitative analysis, we examined mentions of migration, mobility, and porous borders. Often, these codes were mentioned in relation to each other, as with this May 2, 2016 DON alert about the large yellow fever outbreak affecting the DRC and Angola: "Kongo Central province shares a long, porous border with Angola … movement of people between Angola and DRC has [caused] the regular importation of viraemic cases from Angola. The report of Yellow Fever infection in travellers and workers returning from Angola also highlights the risk of international spread of the disease." Porous borders were mentioned in nearly 20% of DON alerts.

The insights from reviewing these alerts add context and motivation to the research aims of this dissertation. The significance of the relationship between human movement and geographic spread of diseases is mentioned throughout the DON archive, yet there is little empirical evidence about the rate of population movement in the DRC. I will now turn to a literature review of the different bodies of scholarship that ground this project.

2. Literature Review

This project draws three distinct bodies of literature to inform the interdisciplinary synthesis of geography, demography, and infectious disease epidemiology. From health geography, I use Meade's Triangle of Disease Ecology as a guiding framework for considering how culture, environment, and behavior shape disease outcomes in space and time. From demography, I employ theories of migration and population mobility to distinguish between different types and routes of population mobility and hypothesize the mechanisms by which human movement and infectious disease outcomes interact in the DRC. From infectious disease research, I draw on the clarity and specificity with which researchers describe core epidemiological concepts such as transmission, surveillance, and diagnosis in my analysis of yellow fever virus in the DRC.

2.1 Health Geography

2.1.1 Health geography, population science, and public health

The subdiscipline of health geography contributes to public health research in two major ways: by offering a methodological toolbox to add spatial and space-time variables into epidemiological and biomedical models of disease, and by contributing theoretical and conceptual frameworks that make explicit the role of geographic context and spatial processes in mediating, causing, or complicating health outcomes. These methodological and theoretical contributions can be traced from the foundations of modern health geography; in fact, they are inherent in the origins of the sub-discipline if we examine the work of Jacques May.

May was a surgeon and public health researcher by training, not a geographer, who found in his work in rural hospitals in Southeast Asia that merely considering "pure" public health and medical variables (i.e. demographic and anthropometric characteristics of his patients) did not fully explain the patterns he was observing in health and disease within the clinic (Emch et al., 2017). He began to create maps of disease distribution and disease vectors to better understand the role of geographic context and the spatial distribution underlying these disease outcomes, and theorized the early foundations of disease ecology by claiming that diseases have organic, inorganic, and social-cultural stimuli (May, 1959). While potentially simplistic by modern standards, May's work laid the groundwork for the contributions that modern health geography makes to population health research: theoretically, where you live matters for your health; methodologically, spatial mapping of diseases can tell us something useful about public health outcomes.

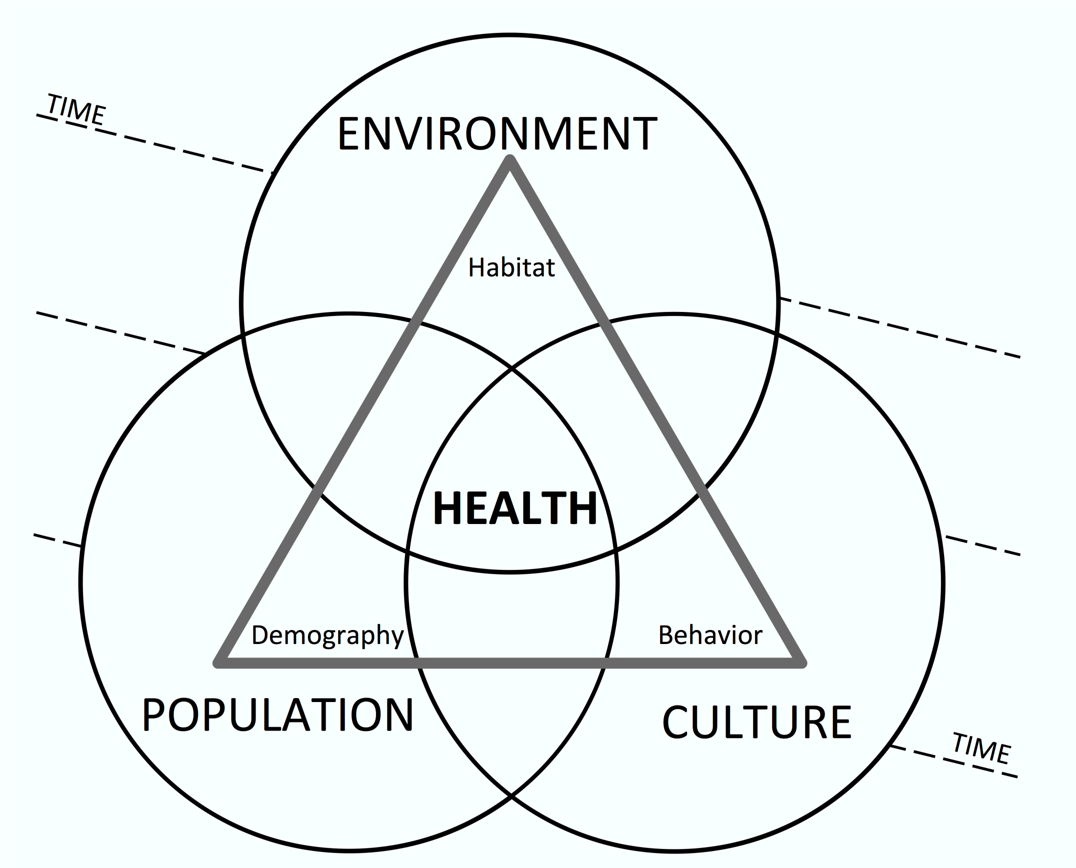
May is thought of as the father of modern medical geography, and his work laid the groundwork for innovations in the field over the subsequent decades by John Hunter and Melinda Meade, among others, who both built on May's initial disease mappings to extend and complicate medical geography and disease ecology into a more nuanced subfield (Hunter & Young, 1971; Meade, 1977). In his chapter in A Companion to Health and Medical Geography (2009), Mayer claims that the subfield of medical geography should really be called epidemiological geography or public health geography, reflecting his view that the contributions of medical geographers are an essential part of epidemiological science and public health practice. He substantiates this claim by pointing out that in any major public health journal, such as the American Journal of Tropical Medicine and Hygiene or the International Journal of Epidemiology, it is common for any given issue to have one or more articles describing studies that include spatial variables, spatial analyses, or other principles from medical geography (Mayer, 2009).

Mayer's observation underscores the methodological contributions that medical geography makes to public health in the modern day. Methods such as spatial statistics (e.g. clustering or hotspot analyses), multilevel modeling, remote sensing, and landscape genetics have their roots in medical geography and have become a core part of epidemiological analysis. Similarly, these spatial tools have also become an essential part of public health practice, as GIS and remote sensing approaches are frequently employed for public health planning of interventions, targeting treatments, and notably for evaluating health care access. This dissertation project builds on the existing interdisciplinary relationships between public health, population science, and health geography to apply spatial tools to questions of population change and disease clustering in the DRC.

2.1.2 Theoretical Frameworks and Disease Ecology

A core contribution that health geography makes to the fields of population health and public health is the introduction of rigorous theoretical frameworks that hypothesize the mechanisms underlying the relationships between behavior, environment, and disease. Returning to Jacques May and Melinda Meade for a moment, May's theory of disease ecology (1959) and Meade's extension of his work into her triangle of disease ecology (1977) have influenced public health research by introducing the concepts of geographic context and the role of space and place into the study of health, wellness, and disease outcomes.

Figure 2. Meade's triangle of disease ecology. Figure adapted from Meade (1977) by Corinna Keeler for publication in Emch, Carrell, and Root (2017).



Meade's triangle of disease ecology frames a disease outcome or state of health in terms of three contexts – population, including genetic factors, gender, age and other demographic characteristic; environment, including natural, social, and built environments; and culture, including cultural beliefs, social organization, individual behaviors, and technology. By introducing the importance of geographic context, May, Meade, and others also opened the door to considering scale effects in public health. The factors encompassed by the umbrella categories of environment, population, and culture each operate at a variety of spatial and temporal scales.

2.2 Demography

2.2.1 Migration, Population Mobility, and Human Movement

Population movement is a conceptually complex demographic category that encompasses a wide variety of human behaviors including migration, commuting, travel, and other forms of mobility (Petersen, 1958). Prothero (1977) theorized the role of human movement in epidemiology, identifying two major types of movement: circulatory and migratory. In this valuable typology, circulatory movement refers to movements where individuals return home after some period of time. Circulatory movement includes anything ranging from short term travel, to migration for seasonal employment, to temporary relocation where the individual returns home (Prothero, 1977). On the other hand, migratory movement refers to a permanent change of residence, with no intention to return to the previous place of residence.

Bell and Ward provide an update to Prothero's framework, distinguishing between temporary mobility and permanent migration (2000) as two dominant types of human movements. Their theorization extends previous work by disentangling the temporal dimension of human movement into three distinct components: duration, seasonality, and frequency (Bell, 2000). Both circulatory and migratory population movements are often theoretically framed and empirically analyzed in terms of push and pull factors (Hristoski & Sotiroski, 2012; Maclin et al., 2017). In this literature, push factors are drivers of out-migration from a certain place, ranging from environmental variables like drought to political variables such as armed conflict (Findlay, 2011; Warner et al., 2010); pull factors are those that incentivize in-migration to a certain place, such as employment opportunities (Hristoski & Sotiroski, 2012).

In the DRC in particular, there is evidence of high levels of population mobility that follows both circulatory and migratory movement patterns. Vlassenroot and Huggins

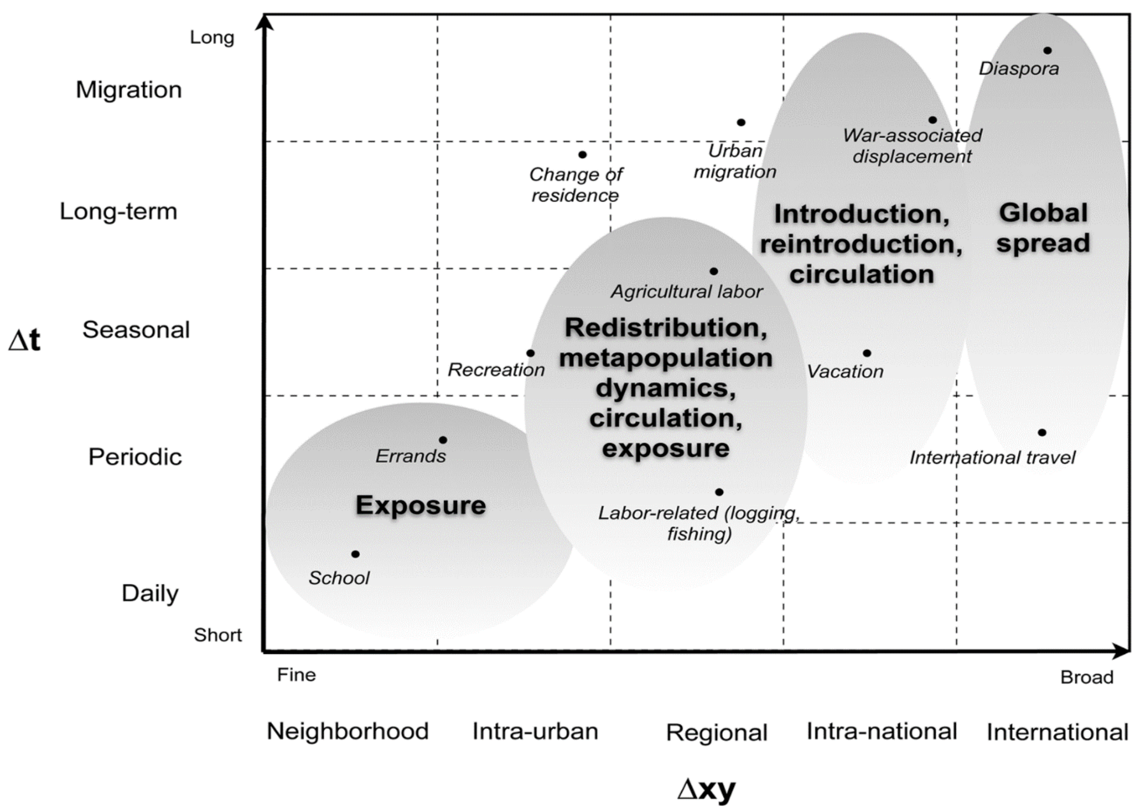
(Vlassenroot & Huggins, 2004a) found that under economic pressure, many young men in Eastern DRC opted to migrate seasonally for work, with rates of economic temporary migration increasing steadily during their study period of 1993-2003. Other studies have found similar patterns in seasonal migration due to participation in farming and mining sectors, especially among men (Geenen, 2014; Maclin et al., 2017). These studies have largely been limited to small study areas or convenience samples (i.e. the employees in a single mine), but have provided valuable data on the typical “push” and “pull” factors for entry into seasonal migration in DRC (Rustad et al., 2016).

In terms of permanent migratory population movement, there is significant evidence of high rates of permanent migration to cities, documented in urbanization rates in the DRC’s capital Kinshasa (Eric et al., 2010; Sanguma, 2015) as well as in Lubumbashi, a significant city in the mining district in the southern part of the country (Andre, 2017). Additionally, over the past 20 years there have been high rates of migration both to and from the DRC of refugees and asylum seekers, as well as large numbers of internally displaced peoples (IDPs) circulating within the country. According to a 2013 estimate contemporaneous with the DHS dataset, there were 2.6 million IDPs in the DRC (UN OCHA, 2013), as well as flows of refugees and returnees between the DRC and neighboring countries of Rwanda, Central African Republic, South Sudan, and Uganda. As the DRC is currently in the midst of a humanitarian crisis recognized by the United Nations and World Health Organization for its severity, this dimension of population mobility has continued to be a significant source of migratory movement in the country (UN OCHA, 2018a).

2.2.2 Relationships Between Human Movement and Infectious Disease

Both circulatory and migratory population movements have significant impacts on the transmission dynamics of infectious diseases, and on epidemiological surveillance efforts and public health interventions. Stoddard and colleagues theorize the role of human movement on vector-borne disease transmission across spatial and temporal scales (Stoddard et al., 2009). They posit that the spatial range of movements, ranging from local to regional to international, intersect with the duration of migration to produce different disease dynamic outcomes. In a key figure from their paper, below, they diagram some of the implications of varying spatial and temporal scales of migration.

Figure 3. Theorized relationships between migration and infectious disease transmission across spatial and temporal scales. Figure is from Stoddard et al. (2009). The x-axis represents spatial scale of population movement, from neighborhood to international. The y-axis represents temporal scale, from daily commuting to permanent resettlement.



This figure illustrates that the spatial *and* temporal scale of population movement intersect to determine the relationship between of population movement and infectious disease dynamics. Activities that occur at the same spatial scale but different temporal scales (such as international travel and diaspora migration) or the same temporal scale but different spatial scales, have different implications for infectious disease transmission.

Human population movements not only impact disease transmission itself, but also the ability of public health practitioners and clinicians to correctly measure case counts and disease rates for planning and intervention purposes. Human population movement can bias disease surveillance efforts by introducing error to both the numerator and the denominator in estimates of disease incidence and prevalence (Buckee et al., 2017). Biases to the numerator occur when cases are misdiagnosed, unreported, or reported in an incorrect location due to travel, circulatory movements, or migratory movements (Buckee et al., 2017; Tatem, 2014). Biases to the denominator are driven by the difficulty of enumerating migrant or mobile populations, which could cause inaccuracies in measuring the underlying population structure, as well as seasonal changes to geographic population distribution that impact the population at risk of seasonal infections (Bharti et al., 2011; Buckee et al., 2017).

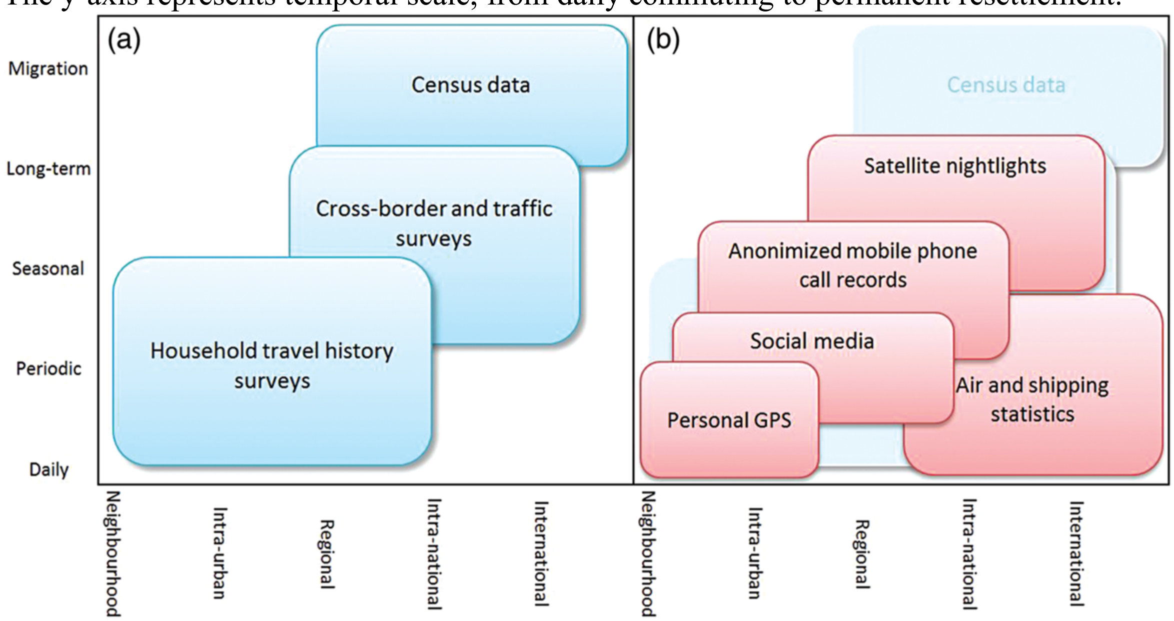
2.2.3 Measuring Human Movement

While the importance of incorporating human movement into estimates of disease rates and models of disease transmission has been clearly demonstrated, data challenges remain an obstacle to actually doing so. Empirical data on human movements and migration patterns is often scarce or incomplete, particularly in the developing world (Tatem, 2014). However, as other sources of data become digitally stored and often openly available, proxy data sources have been used for modeling relationships between human movement and disease transmission.

Beginning over 20 years ago, airline networks and rail networks were broadly available, allowing researchers to use the networked routes and, occasionally, passenger volumes as a proxy for large-scale international or regional movements (Balcan et al., 2009; Brockmann & Helbing, 2013). However, it remains difficult to measure circulatory and migratory population flows that occur via ground transportation, despite the fact that ground transportation is a more common mode of population movement and is hypothesized to drive a much higher share of disease diffusion than migration by plane or train (Pindolia et al., 2012; Tatem, 2014).

The figure below from Tatem (2014) summarizes the data available to quantify human population movements, although notably many of the datasets diagrammed in the figure are not available for many locations nor for many time periods.

Figure 4. Data for measuring human population movements. Figure is from Tatem (2014). The x-axis represents spatial scale of population movement, from neighborhood to international. The y-axis represents temporal scale, from daily commuting to permanent resettlement.



In this diagram, Tatem makes a distinction between "pre-21st century" data sources shown on the left in blue boxes, and "today's" data sources shown on the right in pink boxes. However, as Pindolia et al. (2012) point out, many of the modern data sources outlined above are only used as proxies where census data are not available. Additionally, data sources such as cell phone records and geo-located social media introduce their own questions of geographic accuracy and whether they are adequately representative (Hay et al., 2013). Cell phone data is seen as the best possible movement data available in many parts of the world (Kraemer et al., 2017a; Ruktanonchai et al., 2016), however its utility is limited by its population coverage, the density of cell towers, and finally data access challenges for researchers (Pindolia et al., 2012). Some recent studies have compared census data to cell phone data and found similar relationships between human movement and malaria incidence using both types of data (Garcia et al., 2015; Wesolowski et al., 2013) analysis of benefits of cell phone data versus benefits of census data.

Finally, where none of these empirical or proxy datasets are available, scholars have turned to modeled data. One common approach is to fit a gravity or spatial interaction model on available data in a neighboring country, then apply that model to simulate population movement in the country of interest. This approach has been used to simulate internal migration flows in every malaria-endemic country (Garcia et al., 2015), and to extrapolate population movement in Angola during a Yellow Fever outbreak based on cell phone records from Namibia (Kraemer et al., 2017a).

2.3 Yellow Fever Epidemiology

2.3.1 Yellow Fever Virus Worldwide

Despite the fact that an effective vaccine has existed since 1937, local outbreaks of yellow fever have persisted in parts of Africa, Central America, and South America, with worldwide case counts ranging from 84,000-170,000 per year (Paules & Fauci, 2017). However, in the past two years, two significant outbreaks have recentered yellow fever on the world stage.

A 2015-2016 outbreak in Angola and DRC reached the capital cities of both countries, and more recently a 2017 outbreak in Brazil reached Sao Paolo, by some metrics the most populous city in the Western hemisphere (Vasconcelos & Monath, 2016). Yellow fever is an arbovirus vectored by the *Aedes aegypti* mosquito, the same mosquito that acts as the vector for dengue and zika viruses. The two recent yellow fever outbreaks highlighted a shortage of vaccine (Paules & Fauci, 2017), and prompted a resurgence in yellow fever modeling efforts to simulate worldwide risk areas, model unmet vaccine demand, and examine the epidemiological dynamics of yellow fever outbreaks (Kraemer et al., 2017a; Shearer, Longbottom, et al., 2018; Shearer, Moyes, et al., 2018a).

In the wake of the Angola/DRC outbreak and the Brazil outbreak, several commentaries have focused on the global health risk posed by the possibility of international spread of YFV (Barrett, 2018; Woodall & Yuill, 2016). The 2015-16 outbreak in Angola and the DRC resulted in the first documented occurrence of YFV importation to Asia (Ahmed & Memish, 2017), as 11 confirmed cases in China were traced to the Angola outbreak. This importation sparked concern among researchers, clinicians, and policymakers, as China and much of Asia have densely populated communities of susceptible people, and have been shown to have large populations of the competent mosquito vector *Ae. Agypti* (Ahmed & Memish, 2017; Wasserman et al., 2016). Several recent modelling studies have examined risk of importation of YFV to naïve locations due to travel and migration (N. R. Faria et al., 2018; Wilder-Smith & Leong, 2017), and described possible new risk zones of yellow fever due to changing vector habitat due to climate change (Shearer, Longbottom, et al., 2018).

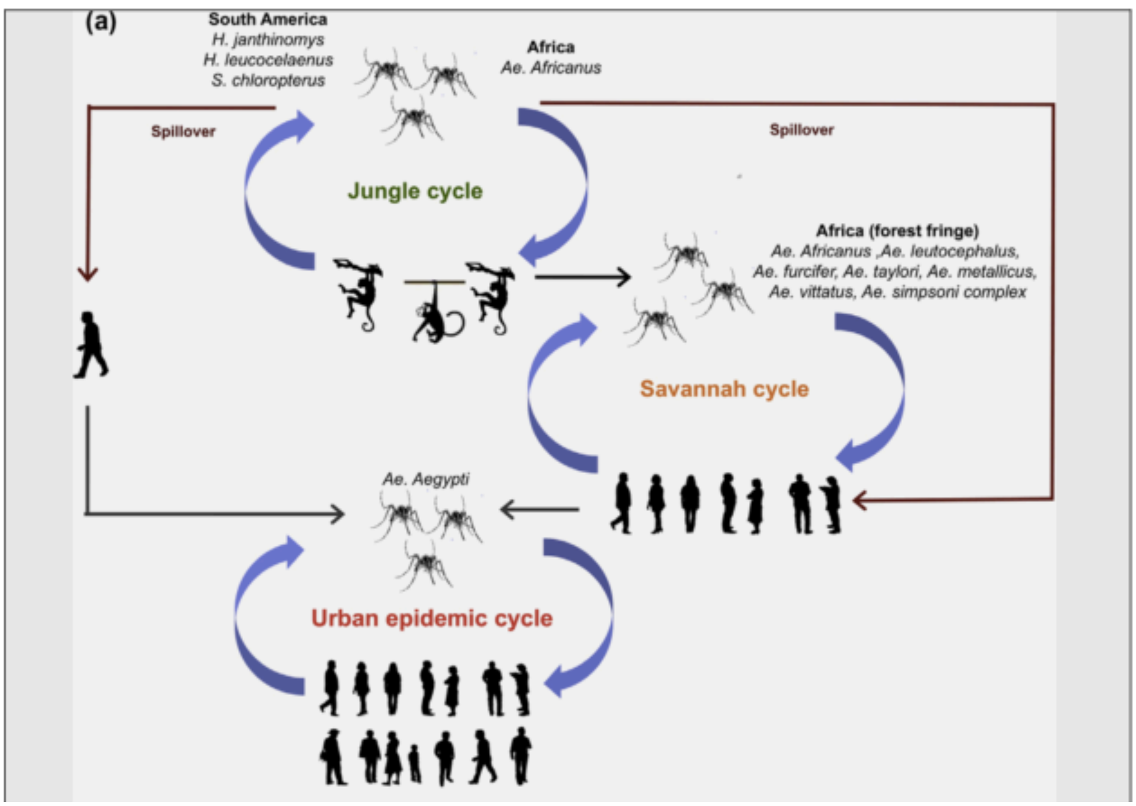
2.3.2 Yellow Fever Virus Transmission

YFV circulation in Africa has been shown to occur in three interrelated transmission cycles: the sylvatic or "jungle" cycle, the savannah or "intermediate" cycle, and the urban or "urban epidemic" cycle (Ahmed & Memish, 2017; Wasserman et al., 2016; Wilder-Smith & Leong, 2017). These interrelated transmission cycles have implications for assessing the drivers of YFV re-emergence, which has been shown to occur more often in areas with recent deforestation and during years with higher than average temperature and rainfall levels (Carrington & Auguste, 2013; Shearer, Longbottom, et al., 2018). Figure 5 shows the three YFV transmission cycles, along with key vector species implicated in each distinct cycle.

The sylvatic, or jungle, transmission cycle occurs between arboreal mosquito species such as *A. africanus* and non-human primates, primarily in forested areas, and allows for the maintenance of an enzootic disease reservoir that periodically spills over into the human population. These spillover events create the savannah or intermediate transmission cycle, which refers transmission between mosquitoes and humans within forests or, more typically, at the forest edge (WHO, 2018). Frequently, this transmission occurs between forest mosquitoes that have experienced habitat disturbances (such as deforestation), bringing them into contact with forest workers or agricultural workers at the forest fringe and therefore resulting in YFV spread within small rural communities (Carrington & Auguste, 2013; Monath & Vasconcelos, 2015).

Typically, the savannah transmission cycle is small-scale and self-limiting, but it can prompt the occurrence of an urban epidemic cycle in cases where infected people are bitten by *Ae. agypti* mosquitoes. This vector species is more anthrophilic than the mosquitos involved in the sylvatic and savannah cycles and is known to live in peri-urban and urban areas (Carrington & Auguste, 2013, Faria et al., 2018). In the past five years, there have been two notable outbreaks of YFV that established urban transmission cycles. In 2015-2016, a YFV outbreak that originated in Angola spread to the DRC, resulting in 16 affected provinces in Angola and 8 affected provinces in the DRC, with confirmed cases in the capital cities of both countries (WHO 2016, Kraemer et al., 2017). Shortly thereafter, a YFV epidemic in Brazil reached the large cities of Sao Paolo and Rio de Janiero, which both have metropolitan populations of over 10,000,000 and were previously understood to be outside the YFV transmission risk zone (N. R. Faria et al., 2018; WHO, 2017b).

Figure 5. Diagram of YFV transmission. The three interrelated transmission cycles – the jungle cycle, the savannah/intermediate cycle, and the urban/epidemic cycle – are shown along with reservoir and vector species. Figure from Carrington & Auguste (2013).



3. Dissertation structure and key contributions

My project draws on the literatures and frameworks described above to explore three interrelated aims. I will describe the primary research questions, data, and findings of the aims below.

3.1 Summary of Chapter 2

The primary aim of Chapter 2 is to describe the spatial population structure of the DRC, given the lack of modern census data described above. The underlying spatial population structure of the DRC has far-reaching implications for infectious disease interventions and control measures. Increasingly, researchers have turned to gridded population datasets such as LandScan and WorldPop to calculate population size for the DRC, however little work has been done evaluating concordance between population estimates obtained from different gridded datasets. This chapter finds that even though national total population estimates are fairly similar between gridded population data sources, the subnational structure of the population at the province and health zone level is drastically different between LandScan and WorldPop data. In order to demonstrate the effect that the selection of population data has on public health intervention efforts, I use the case study of a bednet distribution campaign to illustrate the relevance of understanding the accurate population denominator for implementing disease control measures.

3.2 Summary of Chapter 3

Chapter 3 focuses on human population movement in the DRC. Given the lack of demographic information about the DRC, there is little empirical evidence available about rates of population movement, motivations of Congolese migrants, or geographic patterns in mobility. There are well-documented relationships between population movements and the spread of diseases, however these relationships are often invoked in the case of DRC disease outbreaks without evidence-based links between migration and disease outcomes. The aim of this chapter is to characterize patterns of two distinct types of human movement in the DRC: circulatory mobility, in which people leave home for short- or medium-term durations then return to their home, and resettlement migration, in which people leave without intending to return. I draw on two large cross-sectional population-based surveys that were administered in the DRC in 2012 and 2013 to construct estimates of the rates of each of these types of migration in the DRC, and model the relationship between demographic and employment characteristics and circulatory mobility. I also examine geographic patterns in resettlement migration as well as the dominant motivations for migration among respondents in the survey dataset.

3.3 Summary of Chapter 4

In Chapter 4, I examine the spatial and temporal patterns of yellow fever virus (YFV) occurrence in the DRC. I use data from the Integrated Disease Surveillance and Response system in the DRC that provides week case reports by health zone for the time period 2005-2017, allowing me to examine geographic patterns in YFV suspected cases, temporal trends in YFV cases over the thirteen year study period, and seasonality in YFV cases. I then integrate the data about population mobility and population size uncertainty that I developed in my two previous aims to examine relationships between YFV and demographic factors, and construct statistical models of probability of YFV occurrence using environmental correlates including temperature, precipitation, and deforestation.

CHAPTER 2. THE DEVIL'S IN THE DENOMINATOR: THE RELATIONSHIP BETWEEN POPULATION DATA SELECTION AND UNCERTAINTY IN EPIDEMIOLOGICAL ESTIMATES IN THE DRC

1. Introduction and Literature Review

Epidemiological estimates of infectious disease burdens and at-risk populations inherently rely on underlying datasets that enumerate the size and spatial distribution of a population. However, in many cases, censuses or other comprehensive sources of population data are outdated or unavailable (Linard & Tatem, 2012). The Democratic Republic of the Congo (DRC), which has not had a national census since 1984 (Marivoet & Herdt, 2014), is one such case. With no modern census data, researchers have relied on population data from gridded population models such as the LandScan dataset and the WorldPop dataset (Lloyd, Sorichetta, & Tatem, 2017). Understanding the underlying population distribution of the DRC has significant implications for health research and health service delivery because the country is the site of several recent and ongoing disease outbreaks.

A gridded population model is a spatial demographic approach to population enumeration that defines a grid of uniformly sized grid squares, or pixels, across the spatial extent of the relevant study area, then estimates the number of people living within that grid cell based on a statistical algorithm (Wardrop et al., 2018). As such, these gridded models can be used to define population sizes at a variety of spatial scales by aggregating the estimated number of people living within each grid square in the relevant area, such as national populations, provincial or state populations, and populations within other political or health administrative boundaries. The LandScan gridded population has been modeled worldwide for all years for 2000-2017 and is available at a spatial resolution of 1 km grid squares (Bhaduri et al., 2007). The WorldPop gridded population has been modeled worldwide using country-specific algorithms at two time points, 2010 and 2015, although the WorldPop team has recently released gridded estimates for all of the years between 2000 and 2020 that may have less “tailored geospatial inputs” and methods than the datasets for 2010 and 2015. For all years, these data are available at a spatial resolution of 100m grid squares (Tatem, 2017).

Both datasets are modeled by applying a spatially varying population growth factor to the population count from the most recent national census with administrative boundaries, remotely sensed data, and other supplemental spatial data including road networks, physical geographic features, and settlement locations. However, the specific modeling approach used to dasymetrically assign population to the grid cells differs slightly between the two datasets, though only WorldPop makes their algorithmic data inputs openly available (Lloyd et al., 2017). Previous studies have reviewed available gridded population datasets and compared these data in terms of their spatial and temporal resolution, modeling approaches, and geographic coverage (Wardrop et al. 2018, Leyk et al. 2019, Linard & Tatem 2012, Lloyd et al. 2017). However, to date no studies have empirically evaluated the differences in gridded population data products within a single country in order to better understand how and where the population estimates differ at national and sub-national levels.

Understanding the differences between the different gridded population models is particularly relevant for conducting health research in the DRC for two main reasons. First, since the most recent census occurred over 35 years ago, there are not official census-based population estimates to draw on and researchers often turn to other sources such as non-comprehensive national sources (Marivoet & Herdt, 2014) or the gridded population models described above

(Juran et al., 2018; Kraemer et al., 2017; Nackoney & Williams, 2013). The DRC is an outlier in recency of comprehensive population data, as the United Nations maintains a standard that its member states conduct a census every 10 years (United Nations Population and Housing Programme, 2017) and the DRC is one of just five countries that has not conducted a census since 1985. A quick review of published DRC population estimates illustrates the uncertainty about the population size of the DRC: in 2015 (contemporaneous with the data used in this study), the World Bank estimated the DRC population at 76,244,544, the CIA World Factbook estimated 79,375,136, and the United Nations estimated 77,267,000.

Second, because of the lack of recent census data to inform baseline population estimates and growth factors in these gridded models, the specific modeling differences between the inputs and methods used by WorldPop versus Landscan manifest in significant discrepancies in the modeled spatial population distribution between the two data sources. The differences between these datasets are larger for the DRC than they are for countries in the dataset that have more recent census data available to inform modeling outputs. Therefore, the choice of population data used as a denominator in a given study could strongly impact empirical calculations such as disease burdens, unmet healthcare needs, or expected primary school enrollment that rely on understanding population size and distribution.

This study evaluates the difference between the WorldPop and LandScan datasets within the DRC, with a focus on identifying spatial areas such as provinces and health zones that have particularly large discordances between the two datasets. We demonstrate that these differences between the two datasets in terms of both the size and the spatial distribution of the DRC population are systematic and can be attributed to social and environmental factors. Finally, we use the case study of unmet bednet need in the DRC to illustrate the importance of the underlying population dataset for public health practice and disease control.

2. Data and Methods

2.1 Data

The selection of LandScan and WorldPop data for analysis is described in Appendix I. We obtained the WorldPop datasets from the DRC for 2015 from the Flowminder website. LandScan data were downloaded from the Oak Ridge National Lab website for 2015 in order to temporally match the available WorldPop data. DRC administrative boundaries curated by the United Nations Office for the Coordination of Humanitarian Affairs (UNOCHA) were obtained from the Humanitarian Data Exchange. Selection of administrative boundaries is important because of the gridded nature of the data; small variations in the spatial shape of a province boundary can "reassign" pixels from the gridded population model from one province to another, thus changing the province-level population estimate.

2.2 Population Enumeration Methods

Total populations were enumerated for both WorldPop and LandScan data at three spatial scales: whole-country, province, and health zone level. These spatial scales were selected because they represent meaningful administrative areas to understand population distribution in the DRC: the population of the entire country is a basic measurement used worldwide to describe countries, the provinces of the DRC are the primary unit used to divide the country for both administrative and research purposes, and the health zones of the DRC are the basic areas used for healthcare delivery and distribution of interventions such as bednets and vaccines. The 2015 population for the entire DRC, each DRC province, and each DRC health zone was calculated for each gridded population data input by summing the population of each grid cell that falls within the relevant administrative boundary to generate a total population estimate for the area contained within that specific administrative unit, using the zonal statistics function in ArcGIS 10.5.1.

WorldPop-based population estimates and LandScan-based population estimates were compared at the province and health zone level using correlation coefficients, and areas with the largest magnitude difference in population were identified. We explored the relationships between the discordance in population estimates between the two data sources, and various environmental, economic, and social variables at the province and health zone level that could be relevant for administrative planning, including deforestation, presence of refugees and IDPs, conflict events, and rate of urbanization.

Finally, in order to facilitate a case study application illustrating the effect of population data selection for public health planning, rates of insecticide treated bednet (ITN) ownership were obtained from the 2013-2014 Demographic and Health Survey. By taking the inverse of the ITN ownership rates from the DHS, we generated the proportion of people who did not own a bednet for each province for 2014, or the unmet bednet need. This proportion was applied to the province-level population calculations from both the LandScan and Worldpop datasets, in order to estimate the number of people in need of a bednet as a raw population number instead of a rate. This case study allows us to examine how a bednet distribution campaign might rely on two population characteristics that vary between the LandScan and WorldPop data: the total population size, and the subnational spatial structure of the population.

3. Results

3.1 National population

As shown in Table 1, there were differences in the estimated total population of the DRC for 2015. In order to examine secular trends in DRC population change, we calculated the 2010 population from both population grids as well. WorldPop showed an overall population growth of 12.5% between 2010 and 2015, while LandScan showed an overall population growth of 11.8% between 2010 and 2015. Across both years, WoldPop estimated a larger national population than LandScan.

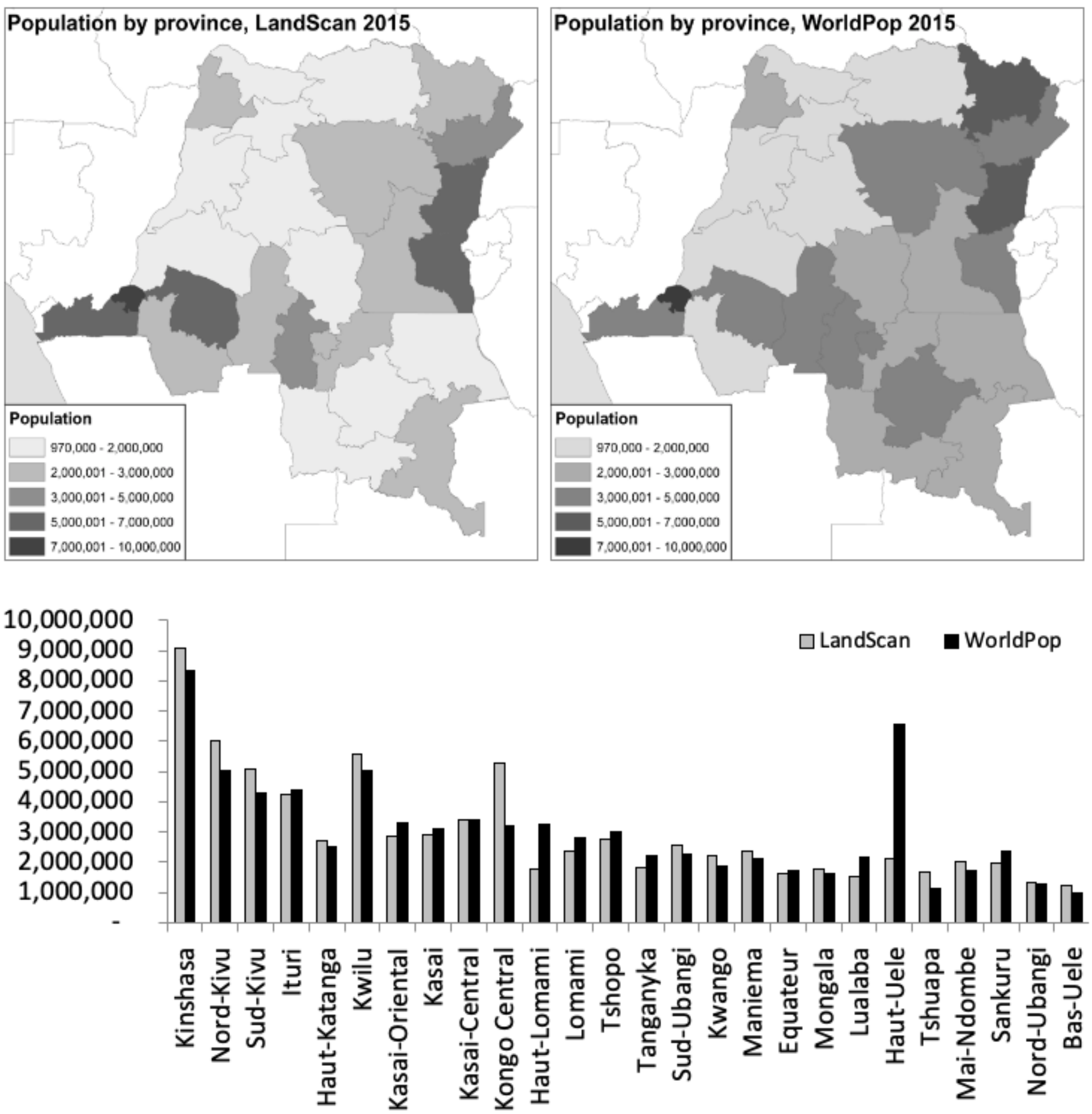
Table 1. National population of the DRC in 2010 and 2015. Populations are calculated from the WorldPop and LandScan gridded population datasets for both time points. WorldPop estimated a higher overall population in both years.

Data source	*2010 DRC population*	*2015 DRC population*
LandScan	70,918,324	79,319,132
WorldPop	72,481,148	81,574,062

3.2 Province-level populations

While the total national population estimated by the two datasets was similar for 2015, province-level populations differed substantially between the LandScan data and WorldPop data. For example, Kinshasa was the most populous province according to both datasets; however, the population of Kinshasa according to the LandScan dataset was 9,076,346 while the population according to WorldPop was 8,351,446 -- a difference of over 700,000. The second and third most populous provinces according to WorldPop were respectively Haut-Uele then Kwilu, while the second and third most populous provinces according to LandScan were North Kivu then Kwilu. Of the 26 provinces in the DRC, 13 (50%) had a higher population according to the LandScan grid than the WorldPop grid. Figure 6, below, shows the spatial structure of the DRC population in 2015 as delineated by both the LandScan and WorldPop gridded population datasets.

Figure 6. Province-specific population comparison between LandScan and WorldPop for 2015. While the national level population is similar between the two datasets, by examining the underlying spatial structure of the population by dataset it is clear that the spatial distribution of the population across the 26 DRC provinces differs notably between these datasets.

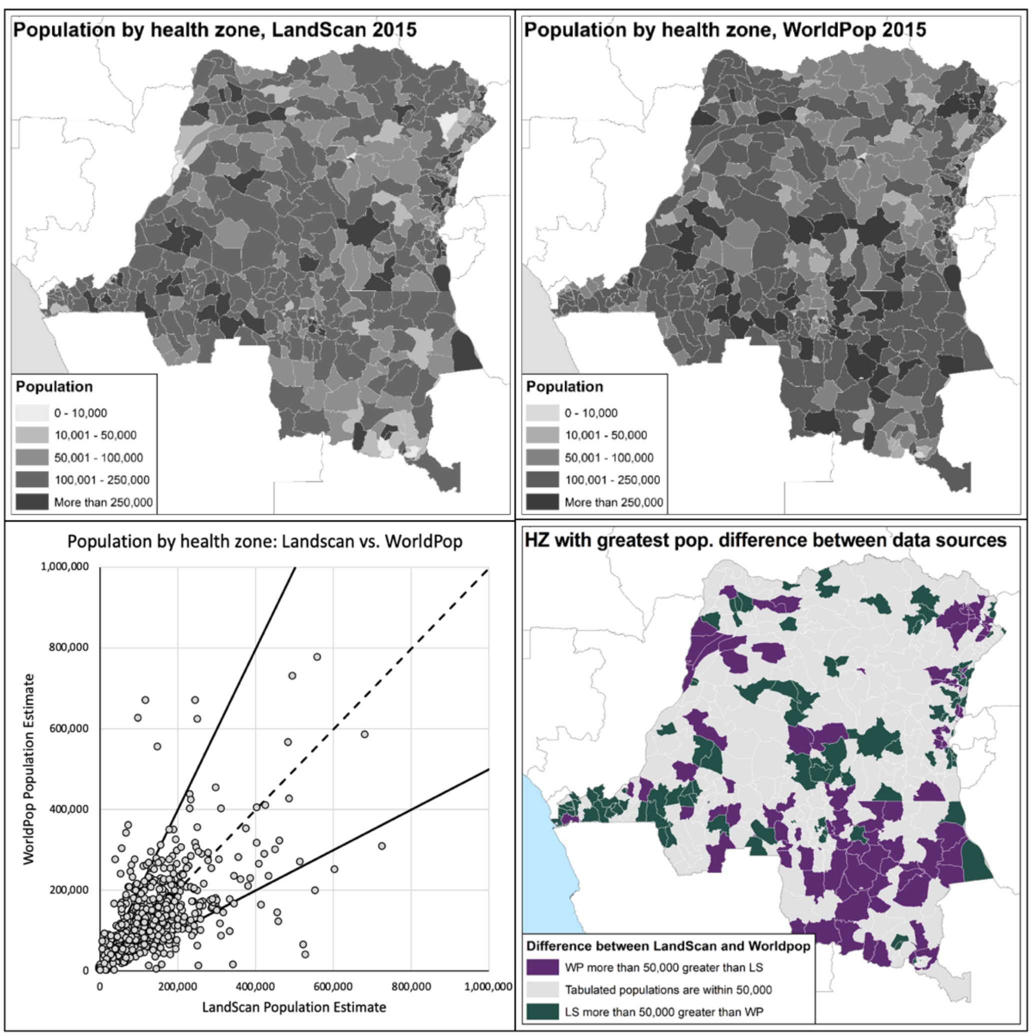


3.3 Health zone-level populations

Of the 515 health zones in DRC in 2015, 111 (21.6%) have a population of at least 50,000 people larger when tabulated using LandScan as compared to WorldPop. Conversely, 108 health zones (21.0%) have a population at least 50,000 people larger when tabulated using

WorldPop as compared to LandScan. Figure 7 shows the health-zone level differences in population in the DRC based on the LandScan and the WorldPop data. By examining differences between the datasets at the health zone scale, patterns in the differences between datasets begin to emerge. The health zones comprising the largest cities in the DRC, including Kinshasa and Lumumbashi, have much larger populations in the LandScan data than the WorldPop data.

Figure 7. Health-zone specific population comparison between LandScan and WorldPop. Panels A and B show the spatial distribution of population by health zone according to LandScan and WorldPop, respectively. Panel C shows the relationship between increasing HZ-level population according to LandScan and the WorldPop estimate for the same health zone, with reference lines indicated for WorldPop = LandScan, WorldPop = 2*LandScan, and WorldPop=0.5*LandScan. Panel D shows the spatial pattern in health zones which have the greatest magnitude population difference between the two datasets.

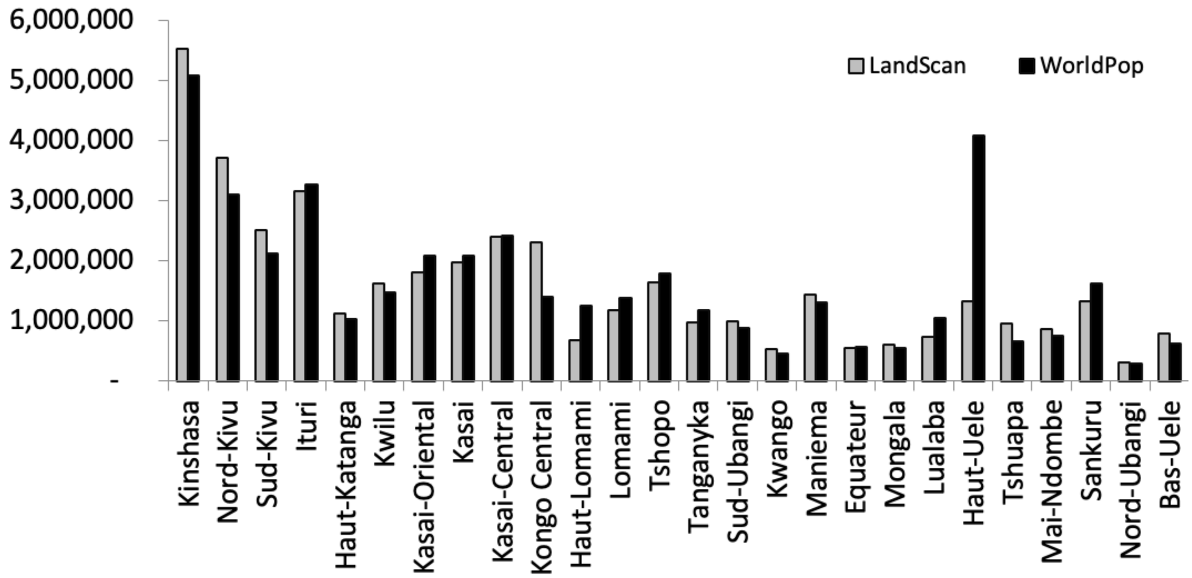


3.4 Case study application of unmet bednet need in the DRC

Accurate population totals are necessary for public health and administrative planning for delivery of services and interventions. The example of unmet bednet need in the DRC demonstrates the importance of having accurate population estimates for provinces and health zones. The burden of malaria remains high in the DRC, and bednet use is a key malaria intervention (WHO World Malaria Report, 2018). Using data from a Demographic and Health Survey (DHS) conducted from 2013-2014 in the DRC, we illustrate the relevance of population data selection for integrated malaria control. While the DHS collects a population representative sample with rich data on bednet use and other malaria preventative measures, these survey data merely allow us to construct a rate of bednet ownership, rather than a *number* of people who own a bednet. Therefore, for actionable public health planning, the bednet ownership rate obtained from the DHS survey must be translated into a raw number of bednets needed per province or health zone in order to inform the allocation of bednets for distribution efforts.

The different province- and health zone-level populations according to the LandScan data versus the WorldPop data result in substantially different estimates of bednet need by province in the DRC, as shown in Figure 8. Based on WorldPop data, and estimated 43,200,316 people lived in households without a bednet in the DRC in 2015, while performing the same calculation using the LandScan data results in a total estimated population without a bednet of 41,153,207. Furthermore, the different population data inputs result in an estimated number of people without a bednet that differs by more than 100,000 people in 19 of the 26 DRC provinces (73.1%) and by more than 500,000 people in four provinces.

Figure 8. Estimation of unmet bednet need by province, based on the selection of underlying population data.



4. Discussion and Conclusion

Because the DRC has not had a census or comprehensive population enumeration since 1984, gridded population data provides necessary denominators for translating prevalences and rates into disease burdens and counts of individuals. Two primary sources of such gridded population data, LandScan and WorldPop, differ in their estimate of the 2015 DRC population by approximately two million people. Furthermore, when examining the population of the subnational administrative units of province and health zone, the spatial structure of the population differs substantially between LandScan and WorldPop. Our findings show that for nearly half of the health zones in the DRC, the selection of input population data between LandScan and WorldPop results in an estimated health-zone level population that differs by more than 50,000 people.

This study uses the example of unmet bednet to illustrate the importance of data selection for public health intervention planning. Additionally, since the DHS is a key source of population and health data in the DRC, a demonstration of the interoperability between DHS data and gridded population data highlights the need for accurate denominators to translate DHS data into policy. While this study was limited to 2015 for data availability reasons, at the time of writing (January 2020), there is an ongoing Ebola outbreak in the northeast provinces of the DRC that is focused in North Kivu and Ituri provinces. The 2015 population of North Kivu as estimated by LandScan versus WorldPop differed by approximately one million people, and the Ituri population differed by approximately 150,000 between the two datasets in 2015. Population numbers are essential for vaccine delivery and vaccine herd immunity threshold calculations, planning for aid delivery including spatial allocation of health care supplies to different health zones, and understanding the underlying reproductive rate and case fatality rate of the epidemic. Uncertainty in population estimates could potentially introduce uncertainty into intervention and control measures for this epidemic as well as future disease outbreaks.

This study has several limitations related to the sensitivity of the analyses. We have no way to validate the true population in the areas with the largest disparity between LandScan and WorldPop and have not conducted any ground-truthing of these data sources. Similarly, this study does not evaluate accuracy or validity of either of these two gridded population datasets. Rather, we examine differences between these data source to highlight the potential impact that selection of gridded population data could have on results and the ensuing recommendations. Finally, we focus on 2015 because it is the most recent year with published data for the DRC from both datasets. The WorldPop Project only creates country-specific algorithms for certain years, and for the DRC the years that have currently published country-specific population models are 2010 and 2015. However, WorldPop does release annual population grids with worldwide population estimates that are often based on continent-wide models of population growth and change. LandScan also releases annual population grids, but since their modeling approach is not openly published, we cannot evaluate whether they have separately calibrated a country-specific model for the DRC in any or all of their available years of data.

We have focused on the DRC as a study area to evaluate differences in gridded population datasets, however this research is applicable beyond just the DRC or central Africa. There are other countries around the world that have not had a comprehensive census in over 30 years, including Somalia, Afghanistan, and Lebanon (United Nations Statistics Division, 2019). In many cases, countries with the longest time gap since a national census are also those experiencing political instability, low rates of infrastructure development, and/or ongoing disease outbreaks. Therefore, these countries may be especially in need of accurate population data. The approach outlined in this study allows national decision makers to quickly compare gridded population data in order to better understand the subnational spatial structure of the population and evaluate the administrative units that have the highest rates of discordance between available population data. While this case study and the example of unmet bednet need highlight the importance of these findings for the DRC, this approach could be used more broadly to understand spatial demography and population changes in countries and regions around the world.

5. From Population Structure to Population Movement

In Chapter 2, I have described the uncertainty in the subnational spatial population structure of the DRC. I now turn to examining demographic and geographic characteristics of human movement in the DRC. These two chapters are thematically linked, as they both deal with fundamental concepts in demography: population size and population movement. As such, they are also linked by the fact that they are each reacting to the lack of available census data in the DRC. Both population size and population movement are parameters that govern the spread of infectious diseases, yet in the absence of contemporary census data, it is necessary to model these components of the DRC's demographic structure using alternative data sources in order to improve infectious disease forecasting. In Chapter 2, I drew on spatially modeled gridded population datasets to create population estimates by province and health zone in the DRC; in Chapter 3, I use large population-representative cross-sectional surveys to construct estimates of human movement in the DRC and map migration flows between provinces.

CHAPTER 3. NATIONWIDE TRENDS IN BOTH CIRCULATORY AND MIGRATORY POPULATION MOVEMENT IN THE DEMOCRATIC REPUBLIC OF CONGO: EVIDENCE FROM TWO CROSS-SECTIONAL STUDIES

1. Introduction

This study characterizes patterns in population movement in the Democratic Republic of the Congo (DRC), given the significance of mobility for infectious disease surveillance, control, and intervention. The DRC is a large country in Sub-Saharan Africa with an estimated population of approximately 83 million people; an estimated 11 million live in the capital and largest city of Kinshasa (CIA, 2019). The DRC is the third most populous country in Africa, has the third highest proportion of urban population on the continent at an estimated 43% in 2017, and is considered to be a rapidly urbanizing country (World Bank, 2018). Additionally, the DRC is the largest country in Sub-Saharan Africa by land area, and it shares a border with nine other countries. These national borders are highly porous (Kabamba, 2010) and international spread of disease across borders has been documented in multiple disease outbreaks (WHO, 2007; WHO, 2010; WHO, 2016).

Prior infectious disease research has pointed to the role of population mobility in spreading and redistributing infectious disease agents and subtypes within the DRC. Lynch & Roper (2011) identified the DRC as a possible corridor of infectious disease transmission between West African countries and East African countries based on the multiple genetically distinct malaria parasite populations found in the DRC, and called for a better understanding of human movement patterns within the DRC. Similarly, molecular surveillance of drug resistant malaria within the DRC has shown ongoing geographic spread of the alleles that confer drug resistance (Aydemir et al., 2018; Deutsch-Feldman et al., 2019). Recently, in a study of HIV subtypes, Faria and colleagues have identified a similar trend, finding high levels of migration of genetic subtypes of HIV-1 both within the DRC and from the DRC to other countries, pointing to high rates of population mobility along those routes (Faria et al., 2019). Despite the well-documented rapid urbanization in the DRC and the fact that the landscape genetics of malaria and HIV point to nationwide population mobility in the DRC, there is a lack of empirical research about human movement in the DRC.

This study leverages available national survey data to examine the demographic and geographic dimensions of human population movements in the DRC. We draw on theories of population mobility to examine two distinct conceptual categories of population movement, circulatory movement and migratory movement. These types of movement have distinct implications for infectious disease transmission (Buckee et al., 2017). By examining the characteristics of population mobility in the DRC, we provide empirical evidence of the high rates of population mobility in the DRC in order to add context to infectious disease surveillance and intervention efforts in the country.

2. Background and Literature Review

Population movement is a complex demographic category that encompasses a wide variety of human behaviors including migration, commuting, travel, and other forms of mobility (Petersen, 1958). Prothero (1977) theorized the role of human movement in epidemiology, identifying two major types of movement: circulatory and migratory. In this typology, circulatory movement refers to movements where individuals return home after some period of time, and can include forms of mobility ranging from short-term travel, to migration for seasonal employment, to temporary relocation where the individual returns home (Prothero, 1977). On the other hand, migratory movement refers to a permanent change of residence, with no intention to return to the previous place of residence. Both circulatory and migratory population movements are often theoretically framed and empirically analyzed in terms of push and pull factors (Hristoski & Sotiroski, 2012; Maclin et al., 2017). In this framework, push factors are drivers of out-migration from a certain place, ranging from environmental variables like drought to political variables such as armed conflict (Findlay, 2011; Warner et al., 2010); pull factors are those that incentivize in-migration to a certain place, such as employment opportunities (Hristoski & Sotiroski, 2012).

Both circulatory and migratory population movements have significant impacts on the transmission dynamics of infectious diseases, and on epidemiological surveillance efforts and public health interventions. Stoddard and colleagues theorize the role of human movement on vector borne disease transmission across spatial and temporal scales (Stoddard et al., 2009). They posit that the spatial range of movements, ranging from local to regional to international, intersect with the duration of migration to produce different outcomes for changing disease dynamics. Human population movements not only impact disease transmission itself, but also the ability of public health practitioners and clinicians to correctly measure case counts and disease rates for planning and intervention purposes. Human population movement can bias disease surveillance efforts by introducing error to both the numerator and the denominator in estimates of disease incidence and prevalence (Buckee et al., 2017). Biases to the numerator occur when cases are misdiagnosed, unreported, or reported in an incorrect location due to either short- or long-term population movements (Buckee et al., 2017; Tatem, 2014). Biases to the denominator are driven by the difficulty of enumerating migrant or mobile populations, which could cause inaccuracies in measuring the underlying population structure, as well as seasonal changes to geographic population distribution that impact the population at risk of seasonal infections (Bharti et al., 2011; Buckee et al., 2017).

There have been no nationwide studies of population movement in the DRC; however, the studies that do exist focus on a small sample, typically limited to one province or region, and document high levels of population mobility that fall into both the circulatory and migratory population movement categories. Vlassenroot & Huggins (2004) found that under economic pressure, many young men in Eastern DRC opted to migrate seasonally for work, with rates of economic temporary migration increasing steadily during their study period of 1993-2003. Other studies have found similar patterns in seasonal migration due to participation in farming and mining sectors, especially among men (Geenen, 2014; Maclin et al., 2017). However, evidence from the DRC has largely been limited to small study areas or convenience samples (i.e. the employees in a single mine, as with (Geenen, 2014)), but have provided valuable data on the typical "push" and "pull" factors for entry into seasonal mobility in DRC (Rustad et al., 2016).

In terms of migratory population movement, there is significant evidence of high rates of permanent migration to cities in the DRC, documented in urbanization rates in the capital city of Kinshasa (Eric et al., 2010; Sanguma, 2015) as well as in Lubumbashi, a large city in the mining district in the southern part of the country (Andre, 2017). Additionally, over the past 20 years there have been high rates of migration both to and from the DRC of refugees and asylum seekers, as well as large numbers of internally displaced peoples (IDPs) circulating within the country. According to a 2013 estimate, there were 2.6 million IDPs in the DRC (UN OCHA, 2013), as well as flows of refugees and returnees between the DRC and neighboring countries of Rwanda, Central African Republic, South Sudan, and Uganda. As the DRC is currently in the midst of a humanitarian crisis recognized by the United Nations and World Health Organization for its severity, this dimension of population mobility has continued to be a significant source of migratory movement in the country (UN OCHA, 2018a).

The anecdotal nature of the available evidence about population movements within the DRC underscores the need for a nationwide examination of the demographic profile of people who are engaging in mobility, the geographic trends in population movements, and the drivers of those movements. Much research on population movements on the African continent and elsewhere draws on census data to elucidate nationwide migration patterns; however, there has not been a census in the DRC in over thirty years (Marivoet & Herdt, 2014). In the absence of available census data, this study draws on existing nationally representative survey data to provide a set of baseline estimates relating to population movements in the DRC.

3. Data and Methods

3.1 Description of data sources

This study uses data from two large cross-sectional population-representative surveys in the DRC to analyze nationwide patterns in circulatory and migratory population movements. The Demographic and Health Survey (DHS) is a cross-sectional population-based survey that collects information on a variety of demographic and health indicators for men, women, and children. It is implemented by the DRC Ministry of Planning in collaboration with the United States Agency for International Development (USAID). The Enquéte 1-2-3 is a household survey on employment in both the formal and informal sector, with a focus on household income and spending, conducted by the DRC Ministry of Statistics. Table 2 describes the two surveys used in this study and the items relevant to population mobility contained within each survey. The two survey questionnaires differ substantially, and many of the demographic and socioeconomic indicators that the DHS captures are not comparably measurable in the Enquéte 1-2-3. However, the Enquéte 1-2-3 contains information about where respondents lived prior to the survey, allowing for spatial analysis of geographic mobility patterns that is not possible based on the survey items included in the 2013-14 DHS.

Given the structure of the survey items that address migration as well as the theoretical basis for distinguishing two major categories of population mobility, circulatory movements and migratory movements were treated as distinct conceptual processes and analyzed separately.

Table 2. Nationally representative cross-sectional surveys in the DRC containing information about human movement included in this study

Survey Name and Year	Total Number of Adults (age 15+)$^{\Psi}$	Questions about population mobility asked of adult survey respondents:
DHS, 2013-14	27483 (8656 men, 18827 women)	*In the last 12 months, have you been away from your home community for more than one month at a time?*
Enquéte 123, 2012	59702 (28906 men, 30796 women)	*Have you always lived in this locality?* *If no, where did you live before you lived here:* • *What type of place did you live in (large city, small city, rural, etc)?* • *If DRC, what territory did you live in?** • *If other country, was it a neighboring country, a different African country, or other?* *If no, why did you come here?*

$^{\Psi}$ The DHS study design includes separate questionnaires for women, men, and households (HH and children's survey questions are asked of women; men are interviewed separately), with inclusion criteria defining adults as age 15-45 for women age 15-65 for men. In contrast the Enquete 123 is a household survey containing people of all ages in a single survey data file; therefore, for this study the DHS age cutoff for adults of 15 was used to define the sample of adults in the Enquete 123. We did not exclude adults older than 65 because we are not directly comparing the samples and wanted to maintain as large a sample size as possible in the Enquete123 survey.

* The territory refers to the Level II administrative unit in the DRC (N=162).

3.2 Methods for circulatory movement

To examine circulatory movement, we used the survey item from the DHS which captured the mobility of participants who had traveled away from their home for an extended time (at least a month) but then returned to their home community by the time of the survey (see Table 2). This variable was treated as a binary outcome for circulatory mobility. Based on qualitative evidence that patterns of mobility may differ between men and women in the DRC

(Maclin et al., 2017), we stratified all analyses by gender in order to capture the distinct characteristics of mobility among women versus among men.

We conducted a descriptive analysis to compare the demographic profile of those with short-term mobility to those with no short-term mobility in the sample, and identified demographic categories with significant differences between mobile and non-mobile respondents using chi-squared tests for categorical demographic attributes such as marital status and urbanicity, and T-tests for quantitative attributes such as age and household size. Based on this descriptive analysis, we identified variables for inclusion in the logistic regression analysis. We fitted logistic regression models stratified by gender using the binary outcome "mobility/no mobility" in order to evaluate the demographic predictors of circulatory mobility among adults in the DRC, and generated odds ratios for each covariate in order to examine the association between demographic characteristics and mobility. Additionally, we conducted logistic regression to examine the association between short-term mobility and employment among men; there were not significant differences in employment status or employment sector between the mobility group and the no mobility groups among women, so employment regression analysis did not include women. Logistic regression models were evaluated using the Akaike information criterion (AIC) statistic, in order to identify the models best describing the demographic predictors of short-term mobility.

The primary spatial sampling unit for the DHS is the sampling cluster. We constructed weighted proportions of circulatory mobility for each sampling cluster (N=536), for overall mobility and gender-stratified mobility. We examined associations between cluster-level mobility and spatial variables including temperature, rainfall, conflict, and deforestation, and found no significant relationships. Additionally, we evaluated the spatial autocorrelation of high-mobility and low-mobility sampling clusters using the Global Moran's I statistic and found no evidence of spatial patterning in cluster-level rates of mobility for men, women, or overall adult mobility. Therefore, spatial results for circulatory mobility are not presented in this chapter.

3.3 Methods for migratory movement

To examine migratory movement in the DRC, we used the survey item from the Enquéte 1-2-3 which asked respondents if they had always lived at the location they were being surveyed in. Respondents who answered "no" to this question were coded as engaging in migratory mobility. Additionally, survey respondents who reported migration were asked follow up questions about how long they had lived at their current residence, their reasons for moving, and where they lived before living in their current location, as described in Table 2 above.

We identified survey items regarding the marital status, educational attainment, age, and urbanicity of Enquéte 1-2-3 respondents that most closely matched the variables in the DHS in order to examine the demographic characteristics of migrants in the DRC, and compared these demographic categories between migrants and non-migrants using chi-squared tests for categorical variables and T-tests for quantitative variables in order to examine which demographic characteristics differed significantly between migrants and non-migrants.

Using the follow-up migration survey items asked only to the respondents who reported migration, we performed descriptive analyses to explore the respondents' timing, spatial patterns, and motivations for migration. We compared the urbanicity of each respondent's current location to the urbanicity of their former location in order to quantify the proportion of rural-to-rural, rural-to-urban, and urban-to-urban migration within the DRC. Respondents who reported migratory movement were asked a follow-up question about their reasons for moving to their current location, with a preset response list of 6 potential reasons plus "other"; we visualized the responses to this question in order to examine involuntary migration, forced migration, and migration for family reasons in the DRC, and compared these reasons between men and women in order to understand potential gendered differences in migration motivations.

Finally, we used the survey item regarding territory that migrants moved from to create spatial data layers in order to examine geographic patterns in migration. There were two variables in the Enquéte 1-2-3 survey that enabled us to examine geographic patterns in migration. First, all surveys were coded with the province of residence for each household (the province is the Level 1 administrative unit of the DRC; N=26), this was coded as the destination location for migrants. Second, respondents who reported migratory movement were asked a follow-up question about which territory they moved from (the territory is the Level 2 administrative unit of the DRC; N-162), this was coded as the origin location for migrants. We created maps of the number of migrants that reported each province as their origin location and the number of migrants that reported each province as their destination location, and used the territory-level information to examine the proportion of reported migration that occurred between two territories within the same province. Finally, we constructed a map of migration flows between each of the 26 provinces in order to empirically examine the nationwide spatial patterns of migratory movement in the DRC.

4. Results

4.1 Results for circulatory movement

4.1.1 Sample characteristics and descriptive findings

There were 27,483 adults surveyed in the 2013-14 DHS in the DRC; of these, a weighted proportion of 9.0% (1688/18827) of women reported engaging in circulatory mobility as measured by the survey item described in Table 2, and a weighted proportion of 16.7%

(1443/8656) of men reported circulatory mobility as measured by the same survey item. Full sample characteristics are described in Appendix II (refer to Table 9 and Table 9).

We examined demographic characteristics of men and women who engaged in short-term mobility (Table 8, Appendix II) and examined differences in the proportions of people in a given demographic category who engaged in mobility versus those who did not. Based on the descriptive analysis, we identified relevant demographic variables to include in model-based analysis. For women, there were significant differences in rurality, educational attainment, wealth index, and marital status between respondents who engaged in circulatory mobility and those who did not ($p <= 0.05$). For men, these four variables also showed significant differences between people with mobility and those with no mobility; additionally, there were significant differences between these groups in age, seasonality of employment, and sector of employment. Key demographic variables associated with mobility for men and/or women are described in full in Appendix II.

4.1.2 Logistic Regression

Results of the demographic model of mobility for men and women are shown below in Table 3. For women, all levels of education were significantly predictive of circulatory mobility compared to no education, while for men only levels of education above primary education were predictive of circulatory mobility compared to no education. Rurality was examined as a four-category variable, with the most urban category treated as the referent group. For both women and men, residing in the most rural category (countryside) had the largest effect on mobility: women living in the countryside were 1.56 (95% CI: 1.26-1.95) more likely to engage in mobility than those living in large cities, while men living in the countryside were 2.26 (95% CI: 1.72-2.97) times more likely to engage in mobility than those living in large cities.

Table 3. Logit models of circulatory mobility and demographic characteristics, stratified by gender. Final demographic model for each gender is detailed below.

Variable	Women Estimate	Women P-Value	Women Odds ratio (95% CI)	Men Estimate	Men P-Value	Men Odds ratio (95% CI)
Intercept	*-3.260*	*0.0000*	*0.04 (0.03, 0.05)*	*-2.005*	*0.0000*	*0.13 (0.08, 0.22)*
Educational Attainment						
no education	REF	REF	REF	REF	REF	REF
incomplete primary	0.350	0.0001	1.42 (1.19, 1.70)	-0.150	0.3756	0.86 (0.62, 1.21)
complete primary	0.353	0.0031	1.42 (1.12, 1.80)	0.007	0.9708	1.01 (0.69, 1.47)
incomplete secondary	0.705	0.0000	2.02 (1.69, 2.43)	0.380	0.0146	1.46 (1.09, 2.00)
complete secondary	0.473	0.0002	1.61 (1.25, 2.06)	0.405	0.0146	1.50 (1.09, 2.09)
higher	0.518	0.0039	1.68 (1.17, 2.37)	0.572	0.0022	1.77 (1.23, 2.57)
Rurality						
capital, large city	REF	REF	REF	REF	REF	REF
small city	0.372	0.0135	1.45 (1.07, 1.94)	0.372	0.0646	1.45 (0.97, 2.13)
town	0.362	0.0009	1.44 (1.16, 1.78)	0.623	0.0000	1.87 (1.41, 2.47)
countryside	0.447	0.0001	1.56 (1.26, 1.95)	0.813	0.0000	2.26 (1.72, 2.97)
Wealth Index						
poorest	REF	REF	REF	REF	REF	REF
poorer	-0.027	0.7618	0.97 (0.82, 1.16)	0.259	0.0080	1.30 (1.07, 1.57)
middle	0.076	0.3744	1.08 (0.91, 1.28)	0.176	0.0706	1.19 (0.99, 1.44)
richer	0.356	0.0001	1.43 (1.20, 1.70)	0.372	0.0003	1.45 (1.19, 1.77)
richest	0.121	0.3429	1.13 (0.88, 1.45)	0.365	0.0200	1.44 (1.06, 1.96)
Marital status						
married	REF	REF	REF	REF	REF	REF
never in union	-0.083	0.2141	0.92 (0.81, 1.05)	-0.808	0.0000	0.45 (0.37, 0.53)
living with partner	0.034	0.6407	1.03 (0.90, 1.19)	-0.155	0.0965	0.86 (0.71, 1.03)
separated	0.433	0.0000	1.54 (1.26, 1.88)	0.184	0.2922	1.20 (0.84, 1.68)
widowed	0.217	0.1994	1.24 (0.88, 1.71)	-0.257	0.4815	0.77 (0.35, 1.50)
divorced	0.610	0.0001	1.84 (1.35, 2.46)	0.617	0.0139	1.85 (1.11, 2.99)
Age (years)	-	-	-	-0.015	0.0000	0.99 (0.98, 0.99)

A five-category wealth index was used to examine the association between socioeconomic status and mobility. Compared to the poorest category, there was a significant positive effect of increased wealth on mobility among men, while the pattern among women was less clear. Men in the two highest-wealth categories (richer and richest) had odds ratios of engaging in mobility of 1.45 (95% CI: 1.19-1.77) and 1.44 (95% CI: 1.06-1.96), respectively, compared to men in the poorest wealth index category. The relationship between mobility and marital status was significant for both men and women. For women, compared to the reference group of those married, people with mobility were more likely to be separated or divorced; there was no significant effect of being single ("never in union"). On the other hand, for men, there were significant relationships with both those who are single and divorced. Compared to married men, single men were less likely to report mobility (odds ratio 0.45; 95% CI 0.37-0.53), while divorced men were more likely to report mobility (odds ratio 1.85; 95% CI 1.11-2.99). Age was a strongly significant predictor of men's mobility, although the effect size was small, and inclusion of age in the men's demographic model improved model fit. Age was not included in the final women's demographic model after evaluation of relative model fitness based on the AIC statistic.

While employment was not significant between women who engaged in mobility and those who did not, employment was significant for men across a variety of survey items including seasonality of employment and sector of employment. We fit logit models of the effect of employment on mobility, and then generated a final men's mobility model combining the final employment model (see Table 11 for model characteristic of men's employment models) and the demographic model described in Table 3 (above). The fully specified final model is described in Table 10 in Appendix II.

Table 4. Employment sector logit model of circulatory mobility and employment among men.

Variable	Estimate	P-Value	Odds ratio (95% CI)
Intercept	-2.374587	0.0000	0.09 (0.08, 0.11)
Employment sector			
no work	REF	REF	REF
agriculture	0.8098828	0.3756	2.25 (1.85, 2.75)
sales	1.2977301	0.9708	3.66 (2.85, 4.71)
professional/technical/managerial	0.7784474	0.0146	2.18 (1.70, 2.80)
army	0.7377915	0.0146	2.09 (1.36, 3.13)
services	0.9160634	0.0022	2.50 (1.97, 3.18)
manual	0.988414	0.0646	2.69 (2.03, 3.54)
clerical	1.0814654	0.0000	2.95 (1.71, 4.90)

The Employment Sector model was selected as the best model of the effect of employment on mobility. This model performed better than the model of employment status and the model of both employment status and employment sector jointly (Table 11, Appendix II). The results of this model are shown in Table 4. Employment in all sectors was significantly predictive of circulatory mobility compared to the referent group of no work, with employment in clerical work and sales showing the highest odds ratios of 2.95 (90% CI: 1.71-4.90) and 3.66 (95% CI: 2.85-4.71) respectively. Agriculture was the largest employment sector among all men sampled in the DHS and was also the largest employment sector among men with mobility. Employment in the agriculture sector was significantly associated with circulatory mobility (odds ratio 2.25; 95% CI: 1.85-2.75).

4.2 Results for migratory movement

4.2.1 Sample characteristics and descriptive findings

There were 59,702 adults surveyed in the 2012 Enquéte 1-2-3 in the DRC; of these, 15,448 (25.88%) responded “no” to the survey item “Have you always lived in this locality?” and were therefore asked follow up questions about their migratory movements in the survey protocol. These 15,448 survey respondents were coded as participating in migratory movement for the purpose of this study. Those who reported migration included 8,401 adult women (27.27% of adult women in the survey) and 7047 adult men (24.38% of adult men in the survey). Full sample characteristics are described in Appendix II (Table 12).

We examined demographic characteristics of men and women who engaged in migration (Table 5), using the available variables that most closely matched the DHS demographic categories analyzed in the circulatory mobility section. Household size and age did not differ significantly between men and woman migrants. However, a greater proportion of women migrants were currently living in a rural area (53.1% of women migrants compared to 45% of men migrants), while a greater proportion of men who reported migration were living in either a city or village at the time of the survey (21.0% and 33.1%, respectively, compared to 18.6% and 21.3% of women). Among respondents who had migrated, there was a higher proportion of men than women in all educational categories above primary, however men reported higher levels of education across both migrant and non-migrant categories (Table 5). Married people comprised the largest share of migrants of both genders, with approximately 55% of both men and women migrants reporting that they were monogamously married at the time of the survey. A larger proportion of women migrants were either divorced (5.7%) or widowed (10.3%) than men migrants (2.2% and 2.0% respectively), while a larger proportion of men migrants were single/never married (28.9%) as compared to women migrants (15.0% single/never married).

Table 5. Demographic characteristics of Enquéte 1-2-3 respondents who reported migratory mobility, stratified by gender.

	Women	Men	All migrants
N	*8401*	*7047*	*15448*
Household size (mean (SD))	6.23 (3.05)	6.24 (3.12)	6.24 (3.08)
Age (mean (SD))	37.08 (15.58)	38.74 (16.10)	37.84 (15.84)
Rurality of current res. (%) **			
City	1562 (18.6)	1482 (21.0)	3044 (19.7)
Village	2378 (28.3)	2333 (33.1)	4711 (30.5)
Rural	4461 (53.1)	3232 (45.9)	7693 (49.8)
Educational attainment (%) **			
None	2634 (31.4)	511 (7.3)	3145 (20.4)
Primary	2589 (30.8)	1466 (20.8)	4055 (26.2)
Secondary	2776 (33.0)	4050 (57.5)	6826 (44.2)
Non-formal Program	26 (0.3)	26 (0.4)	52 (0.3)
University	235 (2.8)	794 (11.3)	1029 (6.7)
Post-University	3 (0.0)	18 (0.3)	21 (0.1)
Professional (INPP)	31 (0.4)	73 (1.0)	104 (0.7)
Other	15 (0.2)	19 (0.3)	34 (0.2)
Missing	92 (1.1)	90 (1.3)	182 (1.2)
Marital status (%) **			
Married- monogamous	4644 (55.3)	3879 (55.0)	8523 (55.2)
Single, never married	1260 (15.0)	2037 (28.9)	3297 (21.3)
Divorced	475 (5.7)	154 (2.2)	629 (4.1)
Married- polygamous	620 (7.4)	461 (6.5)	1081 (7.0)
Common law	526 (6.3)	369 (5.2)	895 (5.8)
Widowed	868 (10.3)	138 (2.0)	1006 (6.5)
Missing	8 (0.1)	9 (0.1)	17 (0.1)

The mean amount of time that migrants had lived in their current location was 13.4 years (interquartile range 3.0-20.0 years). We used this variable to examine the amount of time since the most recent migration, which was right-skewed; over half of the people reporting migration had relocated to their current residence in the previous 10 years.

We examined the relationship between urbanicity of previous residence and urbanicity of current residence (Figure 9). Among all migrants, rural-to-rural migration accounted for 30.3% of migration, rural-to-urban migration accounted for 17.5% of migration, and urban-to-urban migration accounted for 32.0% of migration. Among migrants who were living in a city or village at the time of the survey (N=7543), 4832 (64.1%) reported that they previously lived in a city or village, while 2635 (34.9%) reported that they previously lived in a rural location. Among migrants who were living in a rural location at the time of the survey (N=7537), 4562 (60.5%) reported that they previously lived in a rural location, while 2903 (38.5%) reported that they had lived in a city or village before moving to their current residence. Among respondents who indicated that they had previously lived in a foreign country (N=153), 124 previously lived in a country neighboring the DRC, 15 previously lived in another African country, and 5 had previously lived on another continent (9 had missing data for the follow-up question).

The reasons that men and women who engaged in migration gave for moving are shown in Figure 10. There were notable differences in the motivations that women indicated for migrating versus men; 4483 women (53.4%) reported that following or rejoining family was their primary reason for moving, while 2284 (32.4%) of men gave this reason. On the other hand, 1261 (17.8%) of men reported that their reason for moving was to seek employment compared to 383 (4.6%) of women, and 774 men (11.0%) said that they moved to pursue studies compared to 388 women (4.6%). A similar proportion of women and men reported that they moved due to displacement from war; this reason was given by 431 (5.1%) women and 395 (5.6%) men. All reasons for moving, stratified by gender, are described in Figure 10.

Figure 9. Urbanicity of former residence versus urbanicity of current residence among permanent migrants. Bars represent the number of respondents in the Enquéte 1-2-3 survey who reported migratory mobility and fall into each urbanicity category.

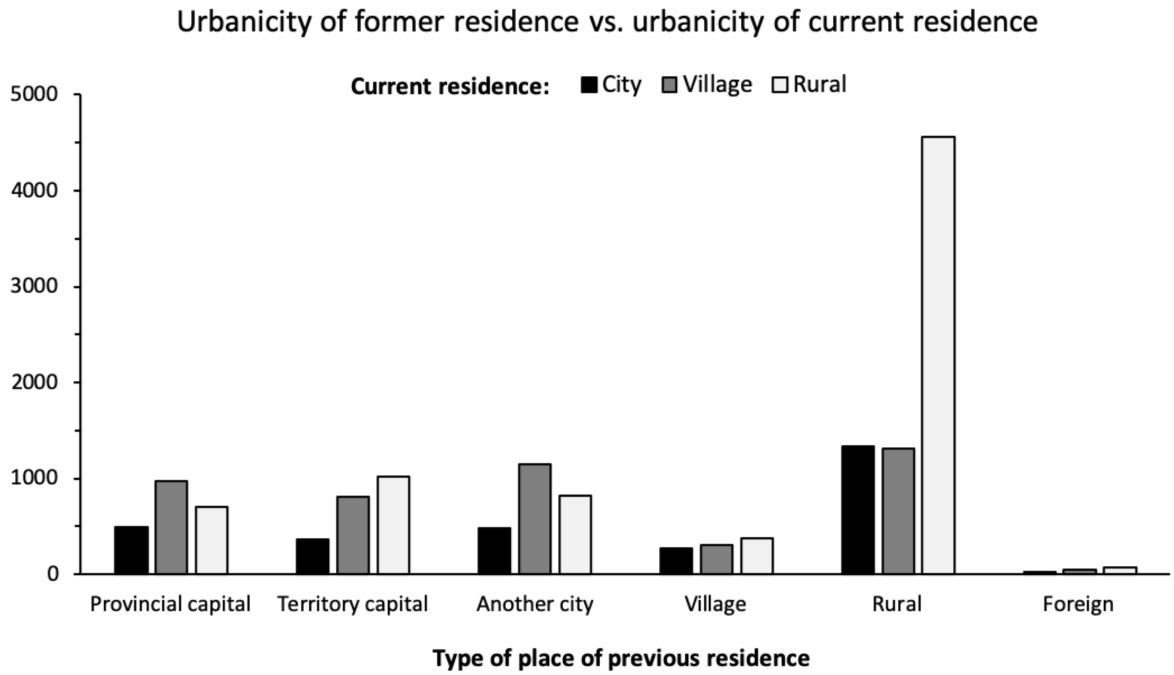
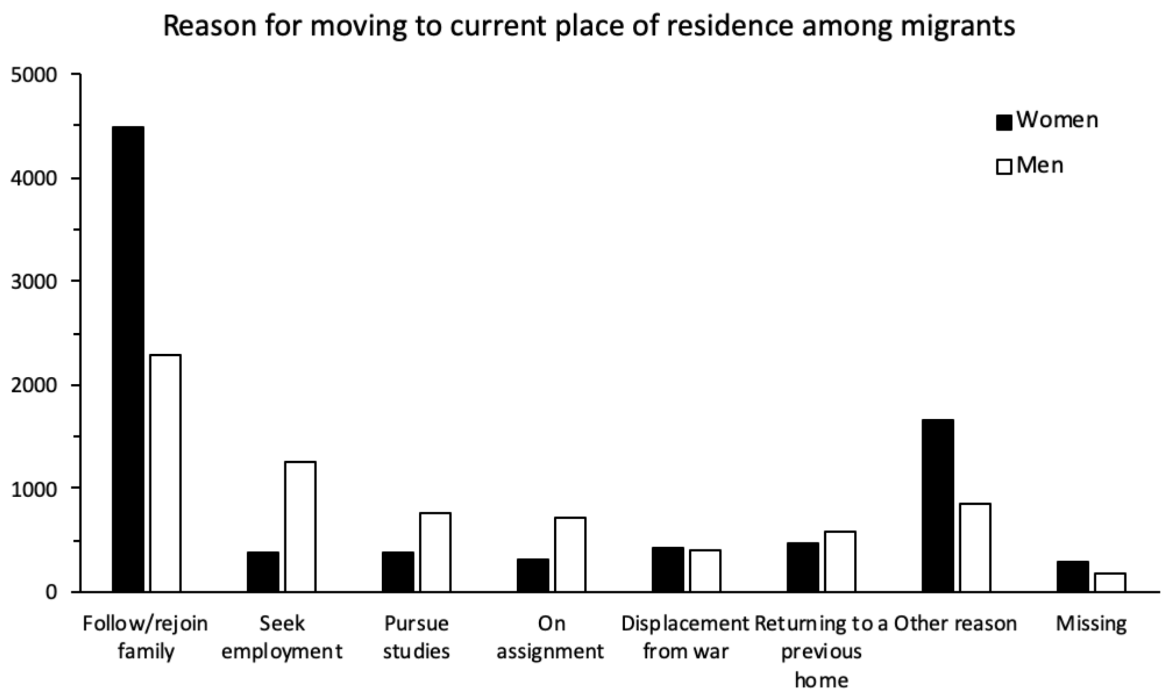


Figure 10. Motivations for permanent migration by gender. Bars correspond to each reason given for moving to current place of residence among Enquéte 1-2-3 respondents with migratory mobility. Black bars indicate the number of women reporting a given reason, and white bars indicate the number of men reporting a given reason.

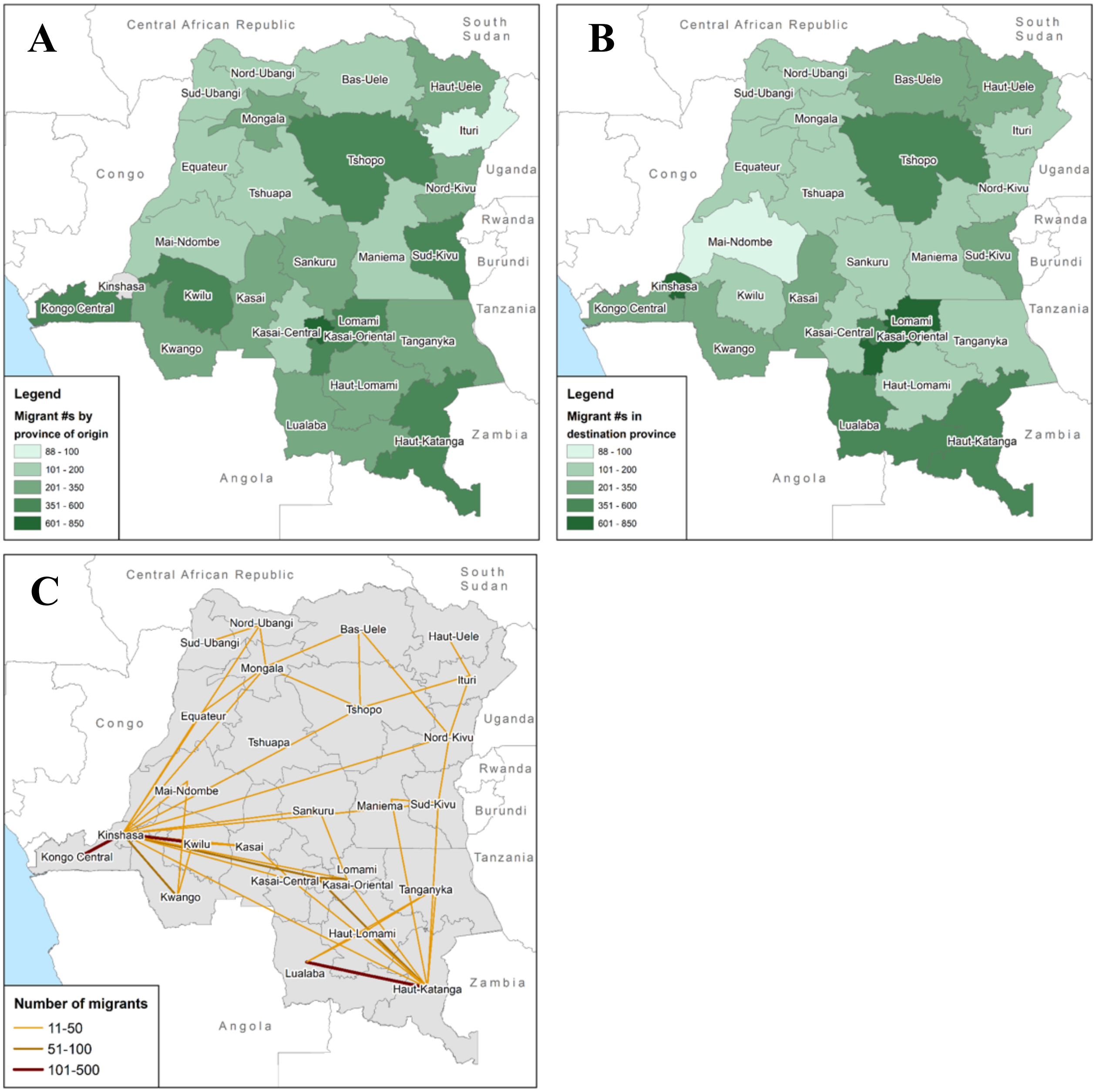


4.2.2 Geographic analysis of migration flows

Migrants reported which territory of the DRC they formerly lived in before moving to their current location of residence. This survey item had high missingness, with only 5441 of the 15448 eligible respondents (35.2%) having data for this survey item. However, since these data provide the only available empirical evidence about nationwide geographic patterns of migration in the DRC, we continued with exploratory analysis to visualize the primary migration flows captured by this survey. For the purpose of data visualization, the territory of origin was aggregated to the province level in order to examine both origin province and destination province of migrants, as well as migration flows (Figure 11). The most frequent destination province for migrants reported in the data was Kinshasa, which contains the capital city (also named Kinshasa), with 13.4% (730/5441) of respondents who had available data about both origin and destination province reporting leaving a different province and moving to Kinshasa.

Within the population of adult migrants with responses about their origin location, 64.5% (3509/5441) reported moving to their current location from a different territory within the same province. The two largest cities in the DRC other than Kinshasa are Lumumbashi and Mbuji-Mayi; 42.6% (194/458) of migrants to Haut-Katanga, the province containing Lumumbashi, reported their origin location as somewhere else within Haut-Katanga province, while 88.3% (257/291) of migrants to Kasai-Oriental province reported an origin location that was also within Kasai-Oriental Province. Migration flows between provinces are shown in Panel A of Figure 11, while the number of migrants in the sample who reported each province as their origin is shown in Panel B and the number of migrants in the sample who reported each province as their destination is shown in Panel C.

Figure 11. Geographic patterns in permanent migration. The maps visualize spatial location of origin provinces (Panel A), spatial location of destination provinces (Panel B), and migration flows (Panel C) among people reporting migratory movement in the 2012 Enquéte 1-2-3.



5. Discussion and Conclusion

We leveraged data from the 2012 Enquéte 1-2-3 survey to examine migratory population movement in the DRC. While much of the existing literature about migration in the DRC focuses on rural-to-urban movements (World Bank, 2018), our findings suggest that rural-to-urban movements do not account for the majority of migratory population movements in the DRC. In fact, urban-to-urban movements and rural-to-rural movements each represented about 30% of the migration documented in these data, while rural-to-urban migration represented 17.5% of the migration in these data. This insight is relevant given the relative lack of demographic information available for the DRC in the absence of a recent national census (Marivoet & De Herdt, 2018), as migratory movements affect administrative planning relating to school enrollment, political representation, vaccination campaigns, and many other civic and government functions (Marivoet & Herdt, 2014). Many administrative functions and resource allocations in the DRC are currently estimated based on a growth factor from the 1984 census, which takes into account fertility rates but may not properly include the spatial redistribution of populations due to migration. Additionally, these findings are relevant in the infectious disease context because much of the focus on migration within infectious disease policy relates to the risks inherent in a certain disease agent or disease outbreak originating in a remote location and reaching a city (WHO, 2016). While our findings do document the existence of rural-urban migration in the DRC in 2012, rural-to-rural migration was twice as frequent, suggesting that there could be possible routes of disease diffusion between disparate rural places in the DRC.

We examined the primary origin locations and destination locations among people coded as migrants in our data, as well as the networked migration flows (Figure 11). Based on the Enquéte 1-2-3, there were two predominant geographic patterns in migration in 2012. First, we observed significant migration from most of the provinces to the capital city of Kinshasa, including provinces like North Kivu and Haut-Katanga that are a significant distance away from Kinshasa in terms of distance and travel time. This pattern aligns with the evidence that Kinshasa is a rapidly growing city with migrants from across the DRC (World Bank, 2018). Second, there were interconnected flows of migration between all of the provinces in the southern part of the country, including the four provinces that were formerly part of Katanga province: Haut-Katanga, Haut-Lomami, Lualaba, and Tangyanika. Notably, these provinces are a primary seat of the mining industry in the DRC (Garrett & Lintzer, 2010; Tsurukawa et al., 2011, DRC Ministry of Mines, 2018). While we do not have the ability within this dataset to connect these population movements specifically to mining, these findings make sense in light of the qualitative evidence from the DRC that participation in the mining industry is a key driver of both circulatory and migratory mobility (Kelly, 2014; Maclin et al., 2017; Rustad et al., 2016).

The survey item that captured migrants' motivations for relocating to their current location allows us to contextualize migration in the DRC in terms of push factors and pull factors. Lack of safety and civil unrest are frequently cited as a dominant "push" factor in the migration literature (Hristoski & Sotiroski, 2012), and are a focus of the international conversation about migration and health in the DRC (UN OCHA, 2018b). About 5% of the people who reported migration gave "displacement from war" as their primary reason for moving. While this is a notable proportion of the migrants in the dataset, other motivations for migrating were more common among both men and women. For women, the dominant motivation for moving was to follow or rejoin family; for men, there seemed to be three "pull" factors that accounted for the majority of the migration observed in the survey, with seeking employment, pursuing studies, or following/rejoining family collectively accounting for 61% of the migration by men in the survey. These estimates could guide the construction of future surveys that better capture individual reasons for migrating in the DRC, in order to elucidate the relationship between time- and space-dependent processes such as land cover change, climate change, and conflict events. A more complete understanding of the push and pull factors governing migration in the DRC would allow for better predictive modeling of migration, which could then be combined with surveillance data to understand the spread of emerging diseases and drug resistant mutations.

The demographic characteristics of people who engaged in short-term mobility also have relevance for infectious disease surveillance and diagnosis. Seasonal or circulatory movements for employment have frequently been linked to disease outbreaks in the DRC, including an outbreak of plague in 2005, yellow fever outbreaks in 2010 and 2016, and an Ebola hemorrhagic fever outbreak in 2018 (WHO 2005, WHO 2010, WHO 2016, WHO 2018); however, little is known about the characteristics of those engaging in circulatory movement. We found that 17% of men engage in circulatory mobility as opposed to 9% of women, and that employment is a significant predictor of mobility in men but not in women. Men with full-time or seasonal employment were more likely to engage in mobility than those who were unemployed, and men who were employed in clerical work, sales, manual labor, or agriculture were the most likely to engage in mobility. Buckee et al. (2017) have posited seasonal population mobility affects epidemiological estimates in several interlocking ways, notably through seasonal misdiagnosis of disease, changing access to medical care based on an individual's changing location, and the effect that changing seasonal population densities have on the denominator in estimating incidences and susceptible populations. By better understanding the characteristics of the mobile population of the DRC, public health professionals, community health workers, and infectious disease researchers can consider the people who are most likely to be missed in surveillance and intervention programs due to mobility, and tailor their programming or health messaging accordingly.

This study has several limitations. We have drawn on available national data sources in order to describe and analyze population movement in the DRC in the absence of other nationwide evidence. However, neither of the two surveys used in this study, the DHS and the Enquéte 1-2-3, were explicitly designed for the analysis of population movements, and the available survey items reflect that. The DHS was designed to collect data related to demographic characteristics and health behavior of respondents, and the one survey item within the questionnaire that relates to population mobility does not contain any geographic information about the destination of travel. The Enquéte 1-2-3 was designed to collect data about livelihoods and participation in formal and informal employment sectors (Marivoet & Herdt, 2014), and the migration-related survey items are in support of that analytic goal. While this survey does contain follow-up questions that allow us to examine migrant's motivations and geographies, there are high levels of missingness in the responses to these follow-up questions among those who reported migration. Additionally, these surveys are from different years (2012 for Enquéte 1-2-3 versus fall 2013-winter 2014 for DHS) and have different sampling structures, so it is not possible to make direct comparisons between them. Finally, while both of these surveys were designed to be nationally representative and both have a large sample size, the sampling frame for each was based on a growth factor applied to the 1984 census in the DRC, and it is not possible to verify whether the survey design is actually geographically or population representative in the absence of a more modern census (Marivoet & De Herdt, 2018).

This is the first nationwide study of population movements in the DRC, to our knowledge. We have analyzed the demographic characteristics of adult men and women who engaged in circulatory mobility as well as the associations between employment and mobility among men, and described the geography of migratory mobility in the DRC in terms of spatial migration flows, rural-urban gradient in migratory movement, and reasons for migration. While this study establishes baseline rates and characteristics using the most complete available data, future work is needed to integrate circulatory mobility and migratory mobility in the same dataset, and to provide information about the geographic dimension of circulatory mobility, which was not possible given the structure of the survey items in past DHS and Enquéte 1-2-3 surveys. However, given the significance of population movements for infectious disease control (Buckee et al., 2017), the characterization of nationwide trends in population movement adds critical context to infectious disease surveillance and control efforts within the country.

6. From movement of people to the spread of disease

Chapter 3 uses cross-sectional data to characterize patterns in both circulatory mobility and migratory mobility in the DRC. No prior work has estimated nationwide rates of population movement in the DRC or created maps of migration flows using data specific to the DRC context as opposed to all of sub-Saharan Africa, and my findings provide evidence for the oft-touted but little-substantiated claim that there are high rates of population movement in the DRC that complicate prevention and control of disease. Chapter 4 is substantively engaged with the spatial and temporal dimensions of one disease in the DRC: yellow fever (YFV). In contrast to the data and approach used in Chapter 3, Chapter 4 uses a spatially and temporally rich dataset of reported YFV cases from a national disease surveillance system to look at patterns of YFV in the DRC across space and time. I then integrate the data about population mobility and population size uncertainty that I developed in my two previous aims to examine relationships between YFV and demographic factors.

CHAPTER 4. GEOGRAPHY AND EPIDEMIOLOGY OF YELLOW FEVER VIRUS IN THE DEMOCRATIC REPUBLIC OF CONGO

1. Introduction

The twenty-first century has brought increased attention to the public health consequences of Arboviruses such as Dengue virus, West Nile virus, Chikungunya virus, Zika virus, and Yellow Fever virus (YFV). Of these viruses, YFV not only has the highest case fatality rate at 25-50% (Barrett, 2018; WHO, 2018), but also is the only disease in this group that can be prevented with a safe, low-cost, and highly effective vaccine (Barrett, 2017; Garske et al., 2014). However, in recent years, vaccination rates have dropped in endemic countries, with estimated coverage rates below 30% in many age groups in countries across Sub-Saharan Africa (Shearer et al., 2018). Understanding the geographic patterns of YFV emergence and endemicity can facilitate earlier identification of outbreaks and improve targeting of vaccination campaigns to the most relevant communities and areas, especially in countries that the WHO considers to be high-risk for YFV epidemics, such as the Democratic Republic of the Congo (DRC) (WHO, 2018). This study examines 13 years of spatially referenced YFV cases in the DRC in order to identify seasonal trends, geographic patterns, and environmental correlates of YFV occurrence in the DRC.

2. Background and Literature Review

YFV circulation in Africa has been shown to occur in three interrelated transmission cycles: the sylvatic or “jungle” cycle, the intermediate or “savannah” cycle, and the urban or “urban epidemic” cycle (Ahmed & Memish, 2017; Wasserman et al., 2016; Wilder-Smith &

Leong, 2017). These interrelated transmission cycles have implications for assessing the drivers of YFV re-emergence, which has been shown to occur more often in areas with recent deforestation and during years with higher than average temperature and rainfall levels (Carrington & Auguste, 2013; Shearer, Longbottom, et al., 2018). The sylvatic transmission cycle occurs between arboreal mosquito species such as *A. africanus* and non-human primates, primarily in forested areas, and allows for the maintenance of an enzootic disease reservoir that periodically spills over into the human population. These spillover events create the intermediate transmission cycle, which refers transmission between mosquitoes and humans within forests or at the forest edge (WHO, 2018). Frequently, this transmission occurs between forest mosquitoes that have experienced habitat disturbances (such as deforestation), bringing them into contact with forest workers or agricultural workers at the forest fringe and therefore resulting in YFV spread within small rural communities (Carrington & Auguste, 2013; Monath & Vasconcelos, 2015).

While the intermediate transmission cycle is typically small-scale and self-limiting, it occasionally prompts the occurrence of an urban epidemic cycle in cases wherein infected people are bitten by *Ae. agypti* mosquitoes. This vector species is more anthrophilic than the mosquitos involved in the sylvatic and savannah cycles and is known to live in peri-urban and urban areas (Carrington & Auguste, 2013, Faria et al., 2018). In the past five years, there have been two notable outbreaks of YFV that established urban epidemic transmission cycles. In 2015-2016, a YFV outbreak that originated in Angola spread to the DRC, resulting in 16 affected provinces in Angola and 8 affected provinces in the DRC, with confirmed cases in the capital cities of both countries (WHO 2016, Kraemer et al., 2017). Shortly thereafter, a YFV epidemic in Brazil reached the large cities of Sao Paolo and Rio de Janiero, which both have metropolitan populations of over 10,000,000 and were previously understood to be outside the YFV transmission risk zone (Faria et al., 2018; WHO, 2017b).

In the wake of these two outbreaks, several commentaries have focused on the global health risk posed by the possibility of international spread of YFV (Barrett, 2018; Woodall & Yuill, 2016). The 2015-16 outbreak in Angola and the DRC resulted in the first documented occurrence of YFV importation to Asia (Ahmed & Memish, 2017), as 11 confirmed cases in China were traced to the Angola outbreak. This importation sparked concern among researchers, clinicians, and policymakers, as China and much of Asia have densely populated communities of susceptible people, and have been shown to have large populations of the competent mosquito vector *Ae. Agypti* (Ahmed & Memish, 2017; Wasserman et al., 2016). Several recent modelling studies have examined risk of importation of YFV to locations with immunologically naïve populations due to travel and migration (Faria et al., 2018; Wilder-Smith & Leong, 2017), and described possible new risk zones of yellow fever due to changing vector habitat due to climate change (Shearer, Longbottom, et al., 2018).

Despite increasing attention to the significance of YFV as a threat to global health, there has been little research about YFV in the DRC. The only previous study that examines geographic patterns in YFV transmission in the DRC outside of a specific epidemic was Willcox et al. (2017), which used serological samples from the cross-sectional 2013 DHS surveys to test for presence of flaviviruses in the DRC, finding low levels of YFV, dengue, and zika present in the serological data. Additionally, two studies have modelled the epidemic behavior of YFV in the DRC during the 2016 epidemic, associated with the 2015-2016 outbreak, finding that proximity to the Angolan border and demographic covariates were significant predictors of whether a district in the DRC was affected by the outbreak (Kraemer et al., 2017; Zhao et al., 2018). Other than these modelling studies, previous research on YFV in the DRC largely focuses on vaccination, examining the effectiveness of vaccination programs and evaluating the immunogenicity conferred by fractional doses of the live-attenuated YFV vaccine (Ahuka-Mundeke et al., 2018; Casey et al., 2019; Wu et al., 2016). Given the potential threat to global public health posed by yellow fever and the lack of geographic research about yellow fever in the DRC, this study describes the epidemiology and geography of yellow fever virus in the DRC. We also examine relationships between YFV occurrence and demographic and environmental factors in order to improve understanding of drivers of YFV transmission in the DRC.

3. Data and Methods

3.1 Data

YFV data were obtained from the Integrated Disease Surveillance and Response (IDSR) system of the DRC. This disease reporting system was established in 2000 by the World Health Organization Regional Office for Africa (AFRO) in order to streamline surveillance at the subnational level for diseases of particular concern for public health response, and to create a clear process for reporting diseases that are notifiable to the WHO under the International Health Regulations of 2005(Gostin et al., 2016; World Health Organization & Centers for Disease Control, 2010). The IDSR provides a standardized mechanism for health zones in the DRC to report cases and deaths suspected to be associated with 24 distinct diseases to the national Ministry of Health; YFV is one such disease.

The IDSR data consist of weekly reports from individual health zones for any suspected cases of any reportable diseases from 2001 to 2017. The case definition for suspected YFV under the IDSR is: “Any person with acute onset of fever, with jaundice appearing within 14 days of onset of the first symptoms” (WHO and CDC, 2010). Figure 21 in Appendix III provides the full standard case definitions for suspected, probable, and confirmed YFV from the IDSR technical reporting manual. A unique observation was a single YFV report for a particular health zone during a particular epidemiological week; these observations could be either YFV-negative (a particular health zone completed the IDSR YFV reporting protocol in a certain week, reporting zero suspected cases), or YFV-positive (a particular health zone completed the IDSR YFV reporting protocol in a certain week, reporting one or more suspected cases). From 2001-2004, the IDSR data had high levels of missingness, likely due to civil conflict and the novelty of the IDSR system (Hoff et al., 2017); therefore, we narrowed our study period from 2005 to 2017. Table 6 describes the number of observations and number of reporting health zones for each year of the study period.

Table 6. Number health zones with IDSR reporting, YFV observations, and YFV cases for each year of the study period.

Year	# of HZ with any IDSR report (for any disease)	# of HZ with any YFV report	# of HZ with at least one YFV case	Total YFV IDSR reports filed	Total YFV cases reported	Total YFV deaths reported
2005	453	100	33	2150	128	5
2006	464	71	33	1816	95	3
2007	461	60	25	1708	204	9
2008	498	82	46	276	118	12
2009	498	44	42	94	123	14
2010	501	42	42	123	286	8
2011	502	35	34	102	197	8
2012	450	95	62	452	350	22
2013	440	97	97	262	435	37
2014	459	115	82	463	418	19
2015	510	114	114	333	446	26
2016	403	193	191	1363	2674	113
2017	512	157	157	487	675	32

3.2 Spatial and Temporal Patterns in YFV Occurrence

In order to verify that the spatial and temporal patterns in YFV cases did not reflect underlying spatial and temporal patterns in overall health zone IDSR reporting, we compared the spatial and temporal distribution of YFV cases across the study period to the spatial and temporal distribution of cholera cases from the same database across the same study period. Cholera was selected as the comparator disease for two reasons. First, it occurred in every year of the data. Second, we hypothesized that in the absence of reporting bias, cholera would show different geographic patterns in case counts, because it has been documented nationwide in the DRC in WHO outbreak reports (WHO, 2015), and its underlying geographic determinants are likely different from YFV: cholera is a water-transmitted disease in which cases often cluster around major waterways (Rebaudet et al., 2013). The results of this sub-analysis are included in Appendix III. There were not similarities between the health zones that reported YFV and the health zones that reported cholera, as measured by number of cases per health zone, number of years with at least one case per health zone, and number of years with at least one IDSR report per health zone (Figure 22).

After validating that the reported YFV cases in our data were not a reflection of underlying disparities in IDSR reporting by health zones, we calculated spatial and temporal parameters of YFV in the DRC. We examined the temporal trend in YFV across the study period by tabulating total YFV cases, total YFV deaths, and the percentage of health zones reporting any cases of YFV, for each year of our data. For the latter measure, we divided the number of health zones with at least one YFV-positive case report in a given year by the number of health zones that made any IDSR report in the same year, in order to account for IDSR reporting missingness, as well as the fact that the DRC created additional health zones during our study period.

Additionally, we examined suspected YFV cases by epidemiological week across the entire study period, in order to assess seasonal patterns in YFV occurrence. We calculated epidemiological week based on the Centers for Disease Control (CDC) MMWR week calendar

(CDC, 2019), which is also used by the WHO, based on the date of each IDSR report. We then aggregated the reported cases from 2005-2017 by epidemiological week, in order to examine within-year seasonality across the study period. Because the temporal pattern of the 2016 cases follows an outbreak curve that is likely driven by the epidemic spread of YFV from Angola rather than seasonal variation in transmission and exposures (Kraemer et al., 2017), we also examined seasonality using cases from all study years expect 2016.

In order to understand spatial patterns of YFV in the DRC, we aggregated observations by health zone. A shapefile of health zone boundaries was acquired from the United Nations Office for the Coordination of Humanitarian Affairs. We calculated and mapped three primary outcome variables at the health zone level. We examined the total number of suspected cases of YFV and suspected deaths due to YFV for each health zone. Additionally, in order to examine stability in YFV transmission across the study period, we examined the number of years that each health zone reported had at least one YFV-positive observation. We examined Moran's I global statistics of spatial autocorrelation for each of these three HZ-level YFV outcomes.

3.3 Relationships between demographic and environmental variables and YFV

For all modeling, we examined the effect of the relevant covariates on two distinct measures of YFV at the health zone level: total reported cases, and total years with at least one YFV-positive case report. We used both of these measures because they showed different spatial patterns in our descriptive analysis, and we hypothesized that they captured two different dimensions of the geography of YFV in the DRC. Years with at least one YFV-positive case report captures sustained transmission across the study period within a health zone and is agnostic to the number of cases per year, allowing us to capture health zones with stable, low-level transmission. On the other hand, total reported cases per health zone reflects the areas with the highest overall burden of yellow fever virus during our study period.

In earlier chapters of this dissertation, I generated demographic characteristics of provinces and health zones in the DRC, which are leveraged here to examine relationships between YFV and demographic characteristics. In Chapter 2, I described methods and outputs from the calculation of population and population density, for each health zone in the DRC, according to two different gridded population data sources. Additionally, I generated the coefficient of variation for each health zone as a measure of uncertainty in total population size given inconsistency in health zone population estimates based on the LandScan versus WorldPop gridded population data (see Chapter 2). In Chapter 3, I described methods for generating province-level tabulations of origin migration and destination migration for each of the provinces in the DRC (N=26). I used these province-level migration tabulations to classify each province of the DRC into "high" and "low" migration provinces, for both origin and destination of migration. The variables relating to population size, population density, uncertainty in population estimates, and province-level migration were used to fit regression models of the relationship between demographic characteristics and YFV.

Given the evidence from studies based in other African countries that precipitation and temperature are linked to yellow fever virus spillover, we examined relationships between these environmental characteristics and both of our YFV outcome measures using OLS regression. Precipitation and temperature data were calculated from the CRU TS3.10 gridded climate dataset (Harris et al., 2014) at the health zone level. Based on the CRU TS3.10 rasters, we generated health-zone level variables for average temperature, temperature variability, average monthly precipitation, and precipitation variability (see Figure 24 in Appendix III).

4. Results

4.1 Temporal patterns in YFV in the DRC

Across the 13 years of our study period, the annual number of reported YFV cases increased slightly over time, with the exception of 2016. The 2016 data point reflects the high burden of disease during the 2016 outbreak of YFV in both the DRC and Angola. There were 6155 total cases of YFV and 308 total deaths due to YFV reported in the DRC IDSR system from 2005-2017. The overall case fatality rate due to suspected YFV was 5.0%. Of the reported cases, 2674 cases and 113 deaths occurred in 2016. Across all years other than 2016, the mean number of cases per year was 290.1 (interquartile range: 126.6 – 422.3) and the mean number of deaths per year was 16.3 (interquartile range: 8.0 – 23.0). The percentage of health zones reporting at least one case of YFV also increased over the 13-year study period. Figure 12 reflects the annual trends in YFV cases and health zone penetration across the study period.

Comparing the first half of the study period to the second half illustrates the increase in reported YFV cases and the increased share of HZ with at least one case. In the first six years of the study period (2005-2010), there were an average of 159 cases per year, and an average of 7.6% of health zones with at least one case of YFV. On the other hand, across the 6 non-outbreak years from 2011-2017 (not including 2016), the average annual number of YFV cases was 420 and an average of 18.9% of health zones reported at least one case. Graphs showing all YFV cases and deaths across the time period are included as Figure 23 in Appendix III. There was not strong evidence of consistent seasonality in reported YFV cases during the 13 years examined in this study. Case counts were fairly consistent within epidemiological weeks between years, as shown in the bottom right of Figure 13.

Figure 12. Reported YFV cases and percentage of health zones with YFV cases, 2005-2017.

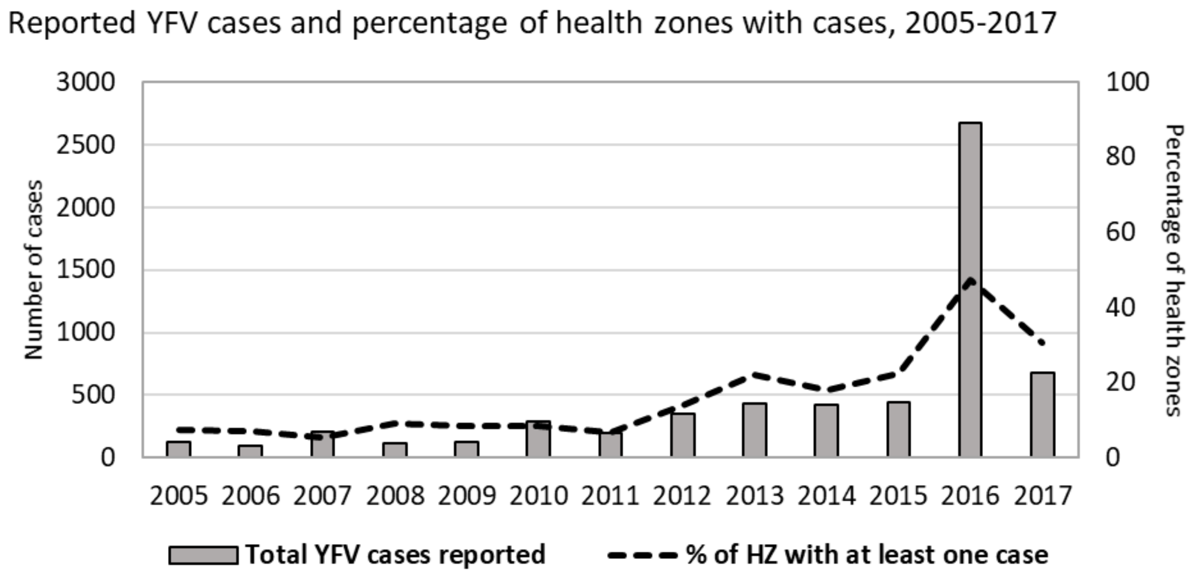
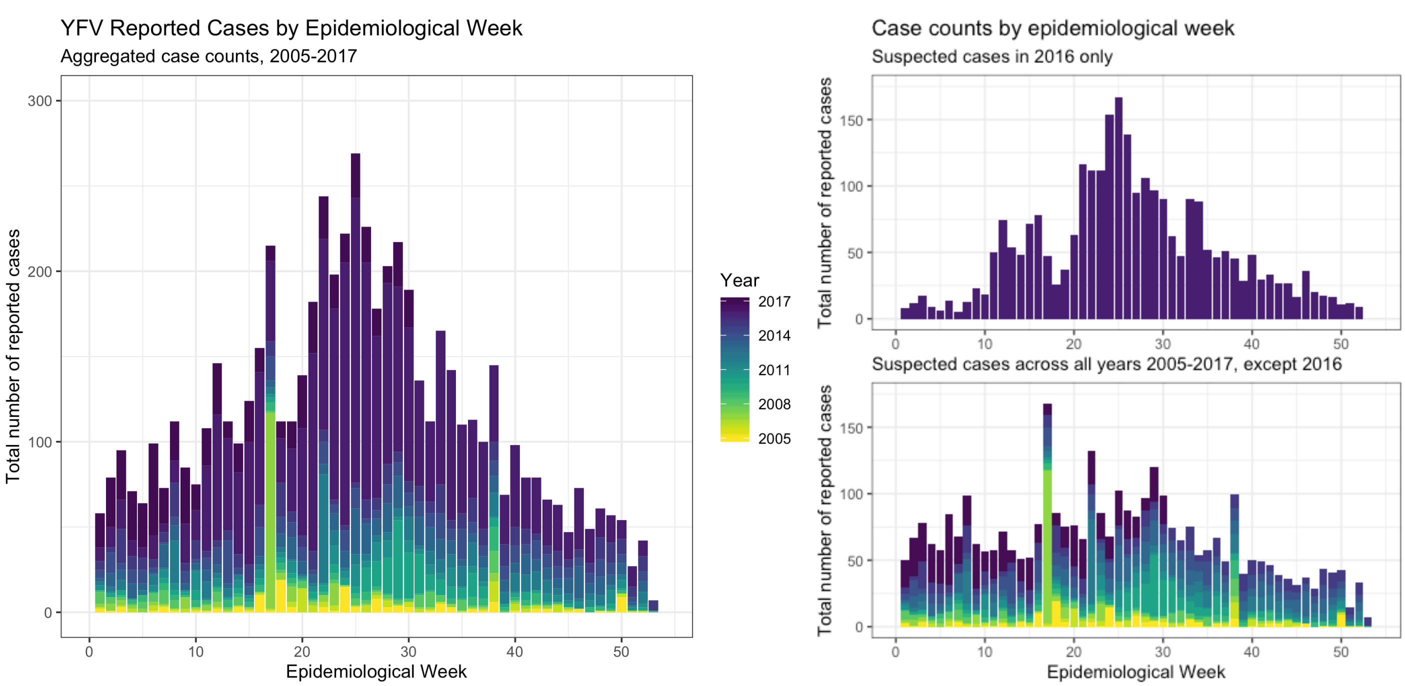


Figure 13. Seasonality in reported YFV cases from 2005-2017. The panel on the left shows cases aggregated by epidemiological week for all years, with the year indicated by the color within the graph. The panel on the right separates out the 2016 cases from the cases in all other study years in order to highlight seasonal patterns that are not driven by the shape of the 2016 epidemic curve.



4.2 Spatial patterns in YFV in the DRC

Maps of total YFV cases and total YFV deaths by health zone are shown in Figure 14. In general, there were three main spatial clusters of cases and deaths: along the Angolan border, in the central-west part of the country in Tshuapa province, and in the northern part of the country is Bas-Uele province. The Moran's I statistic of global spatial autocorrelation in the total cases by health zone was 0.244 (p-value = 0.000) and the Moran's I statistic for total deaths by health zone was 0.19 (p-value = 0.000), indicating significant evidence of strong spatial clustering in both YFV cases and the deaths at the health zone level.

Figure 14. Total YFV cases and total YFV deaths reported by health zone, 2005-2017.

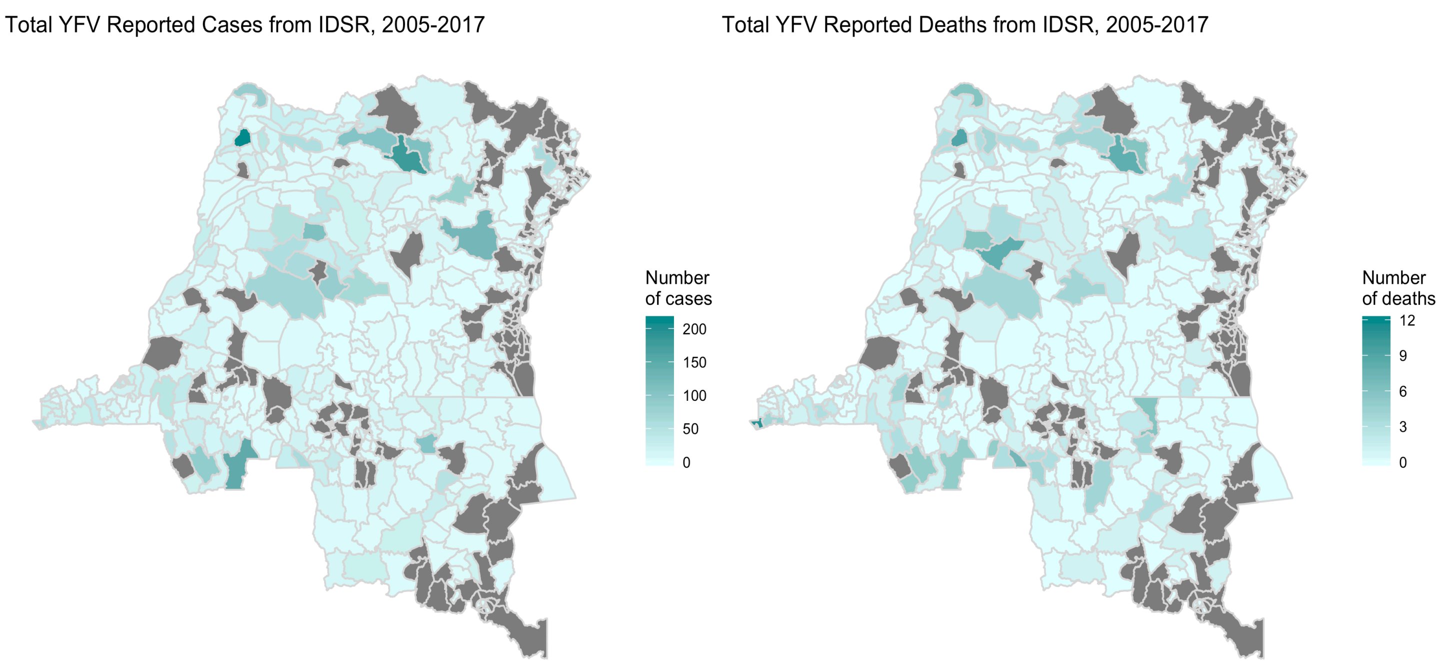
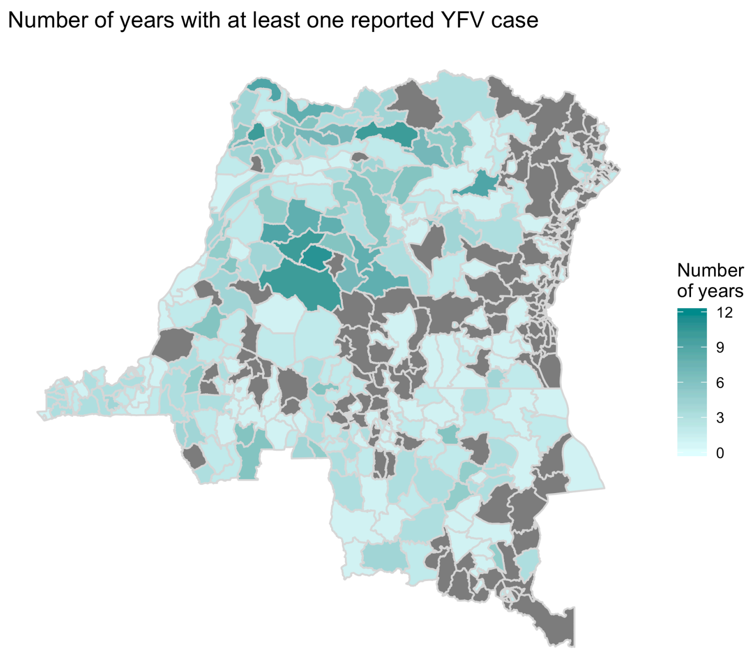


Figure 15 shows the number of years that each health zone reported at least one case of YFV. The range of possible values was 0-13, however there were not any health zones that reported a case of yellow fever in each year of the data. There were five health zones that reported YFV cases in 10 or more of the years during the study period; three of these health zones were located in Thsuapa province (Wema, Monkoto, and Boende). Among all health zones with at least one IDSR YFV report filed, the average number of years with at least one case was 2.68 years.

Figure 15. Number of years with at least one reported YFV case by health zone. Gray areas indicate health zones with no positive or negative IDSR reports for YFV during the study period.



4.3 Relationships between YFV and environmental and demographic variables

Modeling results are shown in Table 7. For the outcome variable of total YFV cases per health zone, the best fit model was the one that incorporated both demographic and environmental variables; both the health zone population coefficient of variation (measure of uncertainty in population size) and the migration destination variable were significantly associated with total YFV cases. The coefficient of variation had a negative effect, indicating that lower levels of uncertainty in population estimates were associated with higher numbers of reported yellow fever cases. There was a positive association between the total number of cases, and being located within a province with higher levels of destination migration. Location within a province with more origin migration was not significantly related to number of YFV cases.

For the outcome variable of number of years with at least of YFV case, the demographic variables were not significant and the best model was the one that incorporated only temperature and precipitation. For both climate variables, the mean and the variability were significantly associated with the number of years a given health zone reported yellow fever. Higher amounts of rainfall and higher temperatures were positively associated with a greater number of years with at least one YFV case. Increased variability in temperature within a health zone had a large positive effect on the number of years with at least one YFV case in that health zone, while increased variability in precipitation levels was negatively associated with years of YFV, although the effect size was small.

Table 7. Model parameters for best fit model for each YFV outcome.

	Outcome = total reported YFV cases		*Outcome = number of years with 1 or more cases*	
Term	**Estimate**	**p-value**	**Estimate**	**p-value**
Precip. mean (mm)	0.29	0.005	0.05	0.000
Precip. variability	0.01	0.869	-0.01	0.024
Temp. mean C	2.74	0.001	0.26	0.000
Temp. variability	12.39	0.002	1.01	0.001
Population variability	-9.89	0.024	*	*
Destination of migration = high	5.91	0.038	*	*

5. Discussion and Conclusion

Despite annual endemic YFV transmission in the DRC, there are few previous studies describing the temporal, seasonal, or geographic patterns of YFV occurrence in the DRC. This study used longitudinal, spatially referenced data from the national disease surveillance system to characterize the geography and epidemiology of YFV in the DRC. We found significant spatial clustering in YFV occurrence, with the highest number of reported cases observed in six contiguous health zones in Tshuapa province, along the border with Angola, and in three contiguous health zones in Bas-Uele province. The case fatality rate (CFR) in across our study period was 5.00%, which is lower than CFRs for YFV reported in the literature (Ahmed & Memish, 2017; WHO, 2018). However, the CFR observed here could be artificially low if there is systematic misclassification of suspected YFV cases, but not misclassification of YFV deaths.

Based on IDSR reports, there is not a strong seasonal dimension of YFV occurrence in the DRC. When aggregating cases across years by epidemiological week in order to examine possible seasonality, the apparent seasonal uptick of YFV between March and August (epidemiological weeks 12-36) can be accounted for by the epidemic curve of the 2016 outbreak. While there is a large body of evidence that mosquito-vectored disease often have strong seasonal patterns (Dery et al., 2010; Mabaso et al., 2005), the evidence about seasonality of malaria in the DRC is weak. Carrell et al. (forthcoming) found minimal seasonal variation in malaria transmission in a longitudinal study. The lack of seasonality in YFV cases in the DRC could be due to the geographically large and varied environment in the country, which precludes defining a single rainy or dry season that spans the entire country.

One limitation of this study is that the YFV surveillance data acquired from the IDSR contain only suspected YFV cases, rather than cases confirmed by laboratory. The case definition for suspected YFV in the IDSR is: "Any person with acute onset of fever, with jaundice appearing within 14 days of onset of the first symptoms" (WHO and CDC, 2010) (the full IDSR protocol for identifying YFV is included in Figure 21 in Appendix III). There is evidence that this case definition for suspected YFV in the DRC is not specific, and studies have shown that some of the YFV cases reported in the DRC IDSR are actually hepatitis family viruses (Makiala-Mandanda et al., 2017), or other arboviruses such as dengue and chikungunya (Makiala-Mandanda et al., 2018). While further research is needed to conclusively establish geographic patterns of confirmed YFV cases in the DRC, the global concern about increased YFV emergence and transmission underscores the need for studies such as this one that leverage available YFV data to better understand the emergence and geographic niche of the disease.

We used two measures of YFV occurrence at the health zone level to examine geographic factors that were correlated with YFV. The total number of suspected cases per health zone was used as a measure of the burden of YFV, while the total number of years with at least one case was used as a measure of the stability of YFV transmission within a health zone. Both of these measures of YFV occurrence showed strong evidence of spatial clustering, indicating that there is an underlying spatial pattern to YFV in the DRC. In particular, the health zones with YFV in the most years could be sites of stable YFV transmission because they are implicated in the intermediate transmission cycle wherein YFV spills into the human population from its enzootic primate reservoir. By examining stability of YFV transmission in addition to overall YFV burden, we have identified health zones that could be targeted for vaccination campaigns. Vaccination in these areas could prevent importation to other health zones that showed a high burden of YFV with lower stability of YFV, thus reducing the overall incidence of YFV in the DRC.

We fit regression models to examine associations between demographic variables, environmental variables, and each of the two YFV outcomes. Higher temperatures, higher precipitation levels, and greater temperature variability were correlated with an increased number of years with at least one case of YFV within a health zone. These findings align with the evidence from a study of YFV seasonality across Africa which found that higher than average rainfall and temperature were strong predictors of a YFV outbreak (Hamlet et al., 2018). While the demographic variables included in our analysis were significantly correlated with the overall burden of YFV, they were not significant for stability of YFV transmission. In particular, we examined relationships between YFV cases and both origin migration and destination migration (see Chapter 3 for further details about the derivation of these variables). While origin migration did not have a significant effect on the number of YFV cases, higher rates of destination migration were correlated with increased numbers of YFV cases reported within a given health zone. This relationship could be an indicator that greater migration flows to certain provinces lead to YFV importation into those health zones, especially along the southern border of the DRC. However, further research is needed to substantiate this relationship and examine causal pathways between YFV transmission and migration, including collecting travel histories among people who report YFV cases in the DRC.

In 2017, in the wake of the YFV outbreak in Angola and the DRC, the World Health Organization launched a new program called the End Yellow Fever Epidemics (EYE) Initiative (WHO, 2017a). This initiative brings together vaccine suppliers, international aid organizations, and national health systems in endemic countries to improve YFV surveillance and streamline YFV outbreak control. A key focus of the program is to identify risk areas to target vaccination campaigns. This study contributes to the larger effort to identify YFV risk areas by examining geographic patterns in YFV occurrence in the DRC and adding to the evidence base about the relationships between YFV and environmental variables.

CHAPTER 5. CONCLUSION AND FUTURE DIRECTIONS

At the time that I am writing this conclusion (February 2019), the global health research and policy community is grappling with the rapidly spreading novel coronavirus 2019-nCoV. This virus is in no way the substantive focus of this dissertation, and I will not over-dwell on it. However, it has been the backdrop against which I have been writing, as social media, online and print journalism, and even most weekly research seminars here at UNC have been heavily focused on the virus' emergence, spread, and political implications. Therefore, it has been on my mind as I consider the whole of my dissertation project and the primary conclusions that have come out of this work. The theme of "moving people, moving pathogens" is strikingly apparent in the global spread of 2019-nCoV, and many scientific minds are already applying themselves to the challenge of identifying spatial and temporal drivers of the spread of the virus over the past 6 weeks in order to predict future spread.

Both mobile populations and infectious disease agents present inherent challenges in identification and measurement. In this dissertation, I reviewed many of the challenges in measuring human movement, and used existing cross-sectional survey data to provide baseline rate estimates of circulatory and migratory population movements in the DRC. Future work could use data with a finer spatial and temporal resolution, such as mobile phone records, to create more robust and dynamic models of human movement in the DRC.

Infectious diseases, on the other hand, present measurement challenges in surveillance and positive diagnosis. In this project, I used surveillance data from the IDSR, a national disease reporting program, which had missing data from at least 10 health zones in every year of our data. As many countries in Africa pivot from the paper-based IDSR to the electronic DHIS-2, there is the potential for increased temporal and spatial completeness in reporting from individual health zones to the DRC Ministry of Health (Angula & Dlodlo, 2018). Even with improved surveillance, confirming cases remains a challenge. The current criteria for a suspected case of YFV is non-specific, and potentially introduces uncertainty in disease incidence estimates due to misclassification of other flaviviruses or hepatitis viruses as YFV.

Finally, both estimates of disease incidence and estimates of population mobility rates require an accurate population denominator to translate proportions into policy-relevant calculations of affected people. While modeled population datasets such as LandScan and WorldPop are increasingly used as population inputs in cases where census data is sparse or outdated, my work shows that the discordance between these two datasets leaves outstanding questions about the underlying spatial structure of the population in the DRC. The recent election in the DRC ushered in a realm of potential political change which could result in the completion of a census for the first time in 36 years. If the census occurs, it will allow government agencies, NGOs, and academic researchers both in the DRC and beyond to better understand the current population in the DRC. As such, the census could fundamentally alter not only our understanding of both population mobility and disease rates in the DRC, as well as the sampling frame and weighting scheme of the surveys from which we derive many of these estimates.

APPENDIX I: SUPPLEMENTAL MATERIALS FOR AIM 1

In this appendix, I document the selection of the two datasets analyzed in this paper. We initially identified four gridded population datasets for inclusion in preliminary analysis, but only two are discussed in the final paper. The four datasets analyzed in the preliminary analysis were the LandScan and WorldPop datasets discussed in the paper, as well as the Gridded Population of the World Data (GPW) and the Global Human Settlement Data (GHS). These four datasets were selected because they all had worldwide coverage and at least one data release for a year after 2010; conveniently, they all had data releases for 2015, so that year worked as a focal time point. The selection of 2015 was driven in part by WorldPop data availability, so it remained the focal time point for the main analysis in the paper.

The figures below illustrate the geographic differences in subnational structure between the datasets (Figure 16), the mean population size with error estimates for each population (Figure 17), and health-zone level measures of correlation between the four datasets. At the health zone level, we calculated a correlation matrix with the correlation coefficient between each of the four datasets (Figure 19). The correlations between the data were low, ranging from 0.26 to 0.54. Finally, Figure 20 shows the coefficient of variation for each health zone. Coefficient of variation was calculated for each health zone by taking the standard deviation of the four population estimates for that health zone, then dividing the standard deviation by the mean population for that health zone across all four datasets, as described by Abdi (2010). The provided us with a measure of uncertainty in health zone population that was standardized against the overall population size.

Based on these preliminary analyses, we moved forward with including only WorldPop and LandScan in the main analysis. The reason for this decision was simple: both GHS and GPW had underlying data issues for the province of Sankuru, with population values of 0 for almost all pixels in the province in both the GHS and GPW gridded population datasets. This anomaly is quite evident in panels C and D of Figure 16, as well as in the map in Figure 20. It is unclear why the population of the Sankuru province is set to 0 in these datasets, but based on this preliminary finding, we excluded them from the full analysis.

Figure 16. Comparison of the spatial population structure by province in the DRC, based on four distinct gridded population raster datasets.

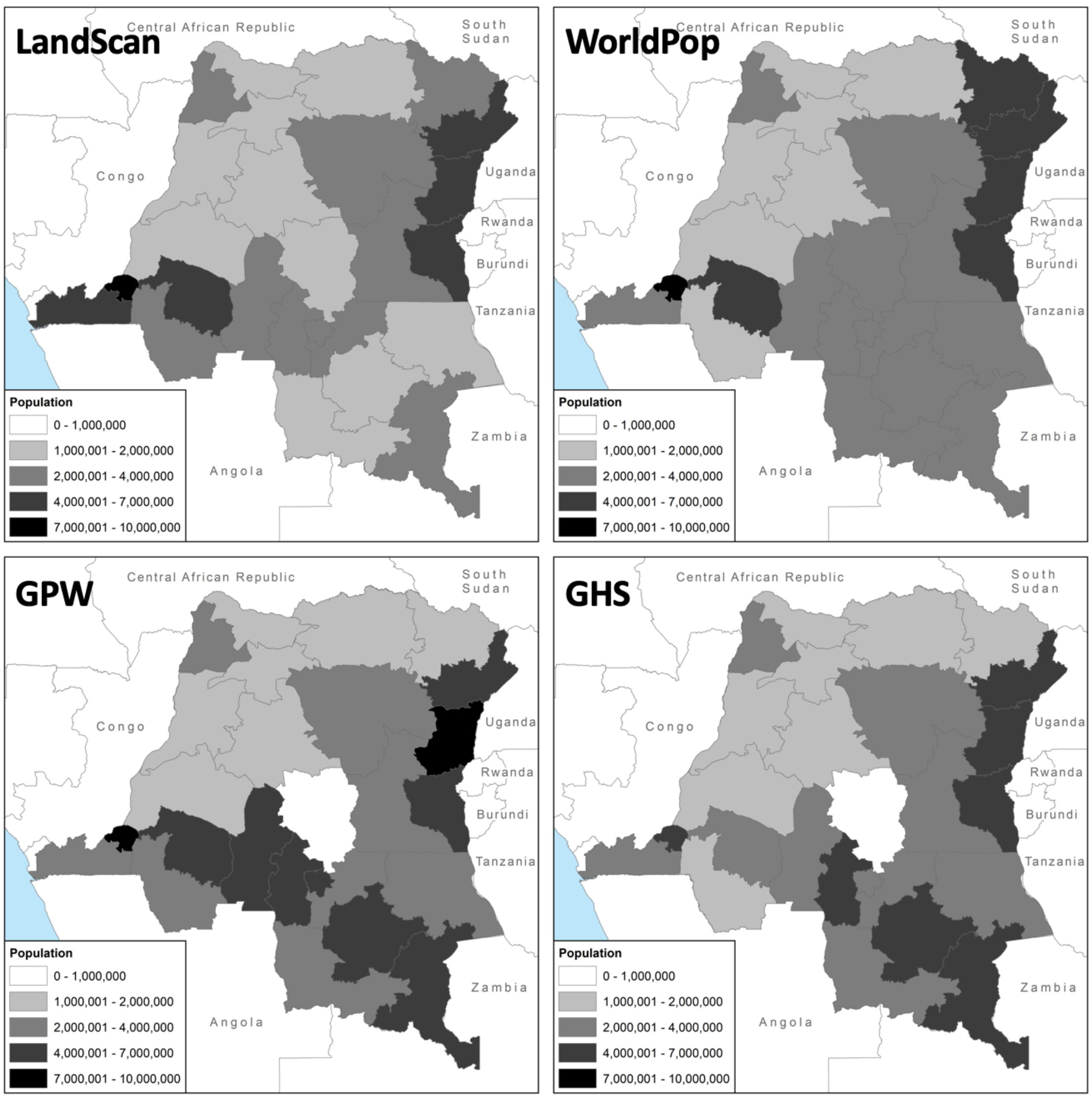


Figure 17. Province population means across all four datasets included in preliminary analysis. Error bars represent 95% confidence interval.

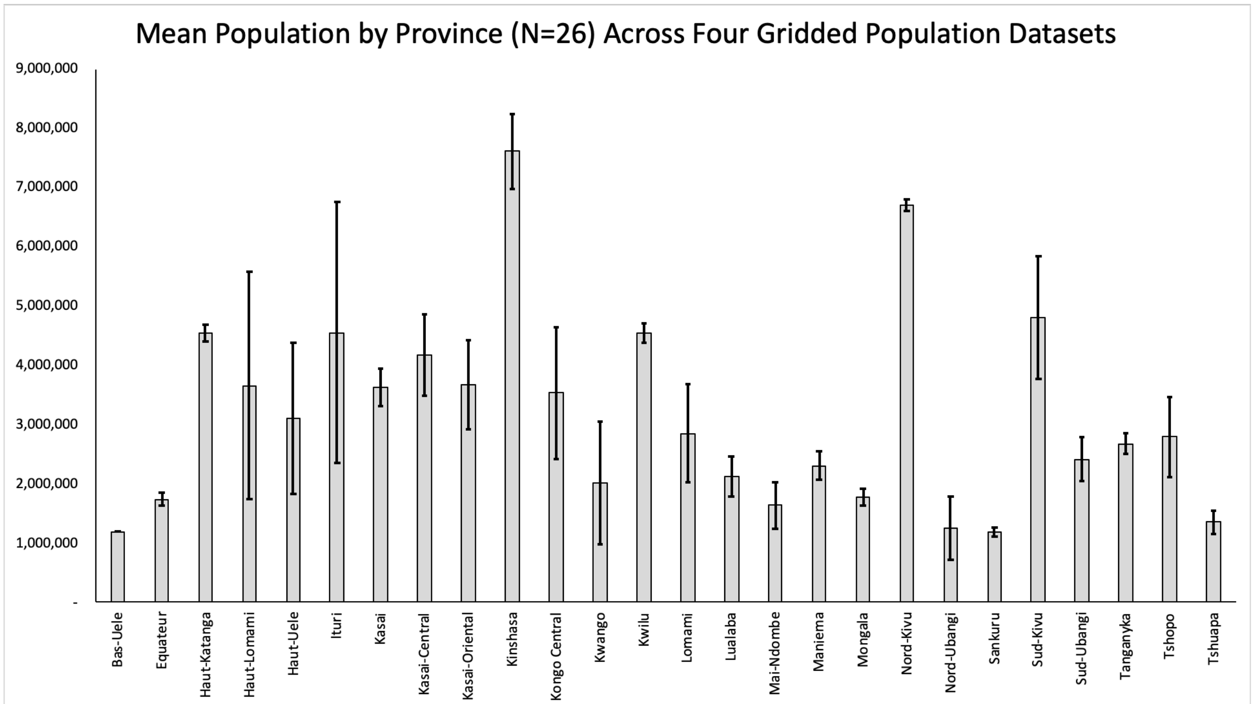


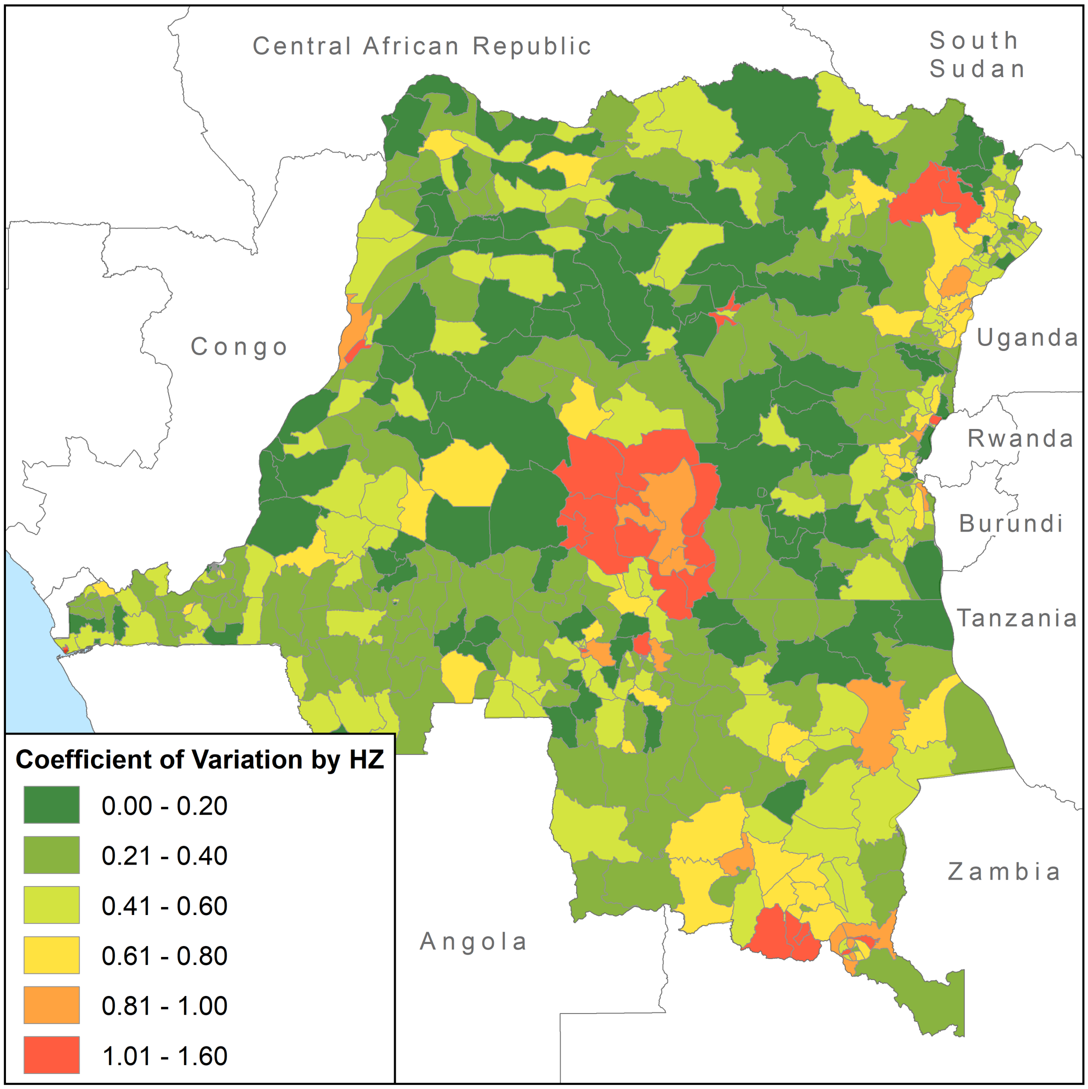
Figure 18. Correlation matrix of similarity in province-level population estimates from four gridded population data sources: LandScan, WorldPop, Gridded Population of the World, and Global Human Settlement.

	Landscan	WorldPop	GPW	GHS
Landscan	x	0.80340	0.69306	0.69297
WorldPop		x	0.60976	0.60996
GPW			x	0.99993
GHS				x
Total Pop.	79,281,062.0	81,561,445.4	89,264,583.8	77,266,292.8

Figure 19. Correlation matrix of similarity in health zone-level population estimates from four gridded population data sources: LandScan, WorldPop, Gridded Population of the World, and Global Human Settlement.

	Landscan	WorldPop	GPW	GHS
Landscan	x	0.31927	0.32168	0.54340
WorldPop		x	0.26453	0.25628
GPW			x	0.51100
GHS				x
Total Pop.	79,281,062.0	81,561,445.4	89,264,583.8	77,266,292.8

Figure 20. Health zone level coefficient of variation across all four datasets. Coefficient of variation was calculated for each health zone by taking the standard deviation of the four population estimates, then dividing the standard deviation by the mean population for that HZ across all four datasets.



APPENDIX II: SUPPLEMENTAL MATERIALS FOR AIM 2

Table 8. Weighted sample characteristics of adults in the 2013-14 DHS Survey in the DRC, stratified by gender.

	Women	Men	All adults
N	18827	8656	27483
Mobility = 1 (N (%))	1688.39 (9.0)	1443.48 (16.7)	3131.87
Current age (mean (SD))	28.09 (9.27)	31.44 (12.23)	59.53
Number of HH members (mean (SD))	6.78 (3.15)	6.67 (3.24)	13.45
Urbanicity (N (%))			
Rural	11602.0 (61.6)	5450.1 (63.0)	17052.1
Urban	7225.0 (38.4)	3205.9 (37.0)	10430.9
Type of place of residence (N (%))			
capital, large city	4433.6 (23.5)	1983.9 (22.9)	6417.5
small city	583.0 (3.1)	244.0 (2.8)	827
town	2208.3 (11.7)	978.1 (11.3)	3186.4
countryside	11602.0 (61.6)	5450.1 (63.0)	17052.1
Educational attainment (N (%))			
no education	2903.0 (15.4)	356.5 (4.1)	3259.5
incomplete primary	5359.6 (28.5)	1380.4 (15.9)	6740
complete primary	1589.8 (8.4)	552.6 (6.4)	2142.4
incomplete secondary	6691.0 (35.5)	4186.0 (48.4)	10877
complete secondary	1595.7 (8.5)	1404.1 (16.2)	2999.8
higher	688.0 (3.7)	776.5 (9.0)	1464.5
Wealth index (N (%))			
poorest	3496.8 (18.6)	1465.7 (16.9)	4962.5
poorer	3588.7 (19.1)	1635.7 (18.9)	5224.4
middle	3510.1 (18.6)	1791.6 (20.7)	5301.7
richer	3654.5 (19.4)	1774.3 (20.5)	5428.8
richest	4576.9 (24.3)	1988.6 (23.0)	6565.5
Total children ever born (mean (SD))	3.05 (2.93)	3.50 (4.31)	6.55
Current marital status (N (%))			
divorced	375.6 (2.0)	83.6 (1.0)	459.2
living with partner	3341.0 (17.7)	999.5 (11.5)	4340.5
married	8754.6 (46.5)	4041.3 (46.7)	12795.9
never in union	4898.8 (26.0)	3247.7 (37.5)	8146.5
no longer living together/separated	1037.0 (5.5)	214.9 (2.5)	1251.9
widowed	420.2 (2.2)	69.0 (0.8)	489.2
Employment (N (%))			
no employment in last 12 months	5219.0 (27.7)	1541.6 (17.8)	6760.6
occasional	3159.2 (16.8)	1122.1 (13.0)	4281.3
seasonal	3979.0 (21.1)	2245.0 (25.9)	6224
all year	6469.8 (34.4)	3747.3 (43.3)	10217.1
Occupation by group (N (%))			
agricultural - employee	66.7 (0.4)	501.2 (5.8)	567.9
agricultural - self employed	7788.0 (41.4)	3169.9 (36.6)	10957.9
army	391.7 (2.1)	201.9 (2.3)	593.6
clerical	65.7 (0.3)	93.5 (1.1)	159.2
not working	5215.9 (27.7)	1539.5 (17.8)	6755.4
others	1.8 (0.0)	1.4 (0.0)	3.2
professional/technical/managerial	531.1 (2.8)	919.9 (10.6)	1451
sales	4236.3 (22.5)	665.7 (7.7)	4902
services	460.1 (2.4)	1021.0 (11.8)	1481.1
skilled manual	2.9 (0.0)	326.5 (3.8)	329.4
unskilled manual	66.8 (0.4)	211.2 (2.4)	278

Table 9. Demographic characteristics of adults in 2013-14 DHS by gender and mobility status. All values are weighted. Significance tests were performed to evaluate which demographic variables varied by mobility status within a gender group. Categorical demographic variables were evaluated for significance using a chi-squared test. Continuous variables were evaluated for significance using a two-way T-test.

	Women				Men			
	All women	**Mobility = 0**	**Mobility = 1**	**p-value**	**All Men**	**Mobility = 0**	**Mobility = 1**	**p-value**
N	*18827*	*17138.61*	*1688.39*		*8656*	*7212.52*	*1443.48*	
Current age (mean (SD))	28.09 (9.27)	28.08 (9.28)	28.20 (9.13)	0.714	31.44 (12.23)	31.27 (12.48)	32.24 (10.88)	0.032
Number of HH members (mean (SD))	6.78 (3.15)	6.79 (3.14)	6.72 (3.17)	0.52	6.67 (3.24)	6.74 (3.28)	6.33 (3.02)	0.001
Type of place of residence (N (%))				0.003				<0.001
capital, large city	4433.6 (23.5)	4100.8 (23.9)	332.8 (19.7)		1983.9 (22.9)	1761.6 (24.4)	222.3 (15.4)	
small city	583.0 (3.1)	519.5 (3.0)	63.5 (3.8)		244.0 (2.8)	208.0 (2.9)	36.0 (2.5)	
town	2208.3 (11.7)	1973.6 (11.5)	234.7 (13.9)		978.1 (11.3)	808.3 (11.2)	169.7 (11.8)	
countryside	11602.0 (61.6)	10544.7 (61.5)	1057.3 (62.6)		5450.1 (63.0)	4434.6 (61.5)	1015.5 (70.4)	
Educational attainment (N (%))				<0.001				0.013
no education	2903.0 (15.4)	2721.6 (15.9)	181.4 (10.7)		356.5 (4.1)	301.3 (4.2)	55.2 (3.8)	
incomplete primary	5359.6 (28.5)	4902.0 (28.6)	457.6 (27.1)		1380.4 (15.9)	1207.3 (16.7)	173.0 (12.0)	
complete primary	1589.8 (8.4)	1454.7 (8.5)	135.1 (8.0)		552.6 (6.4)	471.7 (6.5)	80.9 (5.6)	
incomplete secondary	6691.0 (35.5)	5959.0 (34.8)	732.0 (43.4)		4186.0 (48.4)	3450.3 (47.8)	735.7 (51.0)	
complete secondary	1595.7 (8.5)	1463.5 (8.5)	132.1 (7.8)		1404.1 (16.2)	1135.7 (15.7)	268.4 (18.6)	
higher	688.0 (3.7)	637.8 (3.7)	50.1 (3.0)		776.5 (9.0)	646.2 (9.0)	130.4 (9.0)	
Wealth index (N (%))				<0.001				<0.001
poorest	3496.8 (18.6)	3210.5 (18.7)	286.3 (17.0)		1465.7 (16.9)	1242.9 (17.2)	222.9 (15.4)	
poorer	3588.7 (19.1)	3298.7 (19.2)	290.1 (17.2)		1635.7 (18.9)	1317.7 (18.3)	318.1 (22.0)	
middle	3510.1 (18.6)	3185.8 (18.6)	324.3 (19.2)		1791.6 (20.7)	1467.6 (20.3)	324.0 (22.4)	
richer	3654.5 (19.4)	3226.7 (18.8)	427.8 (25.3)		1774.3 (20.5)	1440.5 (20.0)	333.8 (23.1)	
richest	4576.9 (24.3)	4217.0 (24.6)	359.9 (21.3)		1988.6 (23.0)	1743.9 (24.2)	244.7 (17.0)	
Current marital status (N (%))				0.002				<0.001
divorced	375.6 (2.0)	321.3 (1.9)	54.3 (3.2)		83.6 (1.0)	60.0 (0.8)	23.6 (1.6)	
living with partner	3341.0 (17.7)	3041.2 (17.7)	299.8 (17.8)		999.5 (11.5)	819.7 (11.4)	179.7 (12.5)	
married	8754.6 (46.5)	8016.3 (46.8)	738.3 (43.7)		4041.3 (46.7)	3248.9 (45.0)	792.4 (54.9)	
never in union	4898.8 (26.0)	4477.5 (26.1)	421.2 (24.9)		3247.7 (37.5)	2855.2 (39.6)	392.6 (27.2)	
separated	1037.0 (5.5)	903.8 (5.3)	133.1 (7.9)		214.9 (2.5)	168.6 (2.3)	46.3 (3.2)	
widowed	420.2 (2.2)	378.5 (2.2)	41.7 (2.5)		69.0 (0.8)	60.1 (0.8)	8.9 (0.6)	
Employment (N (%))				0.208				<0.001
no employment last 12 months	5219.0 (27.7)	4784.9 (27.9)	434.1 (25.7)		1541.6 (17.8)	1410.6 (19.6)	131.1 (9.1)	
occasional	3159.2 (16.8)	2834.8 (16.5)	324.5 (19.2)		1122.1 (13.0)	915.1 (12.7)	207.0 (14.3)	
seasonal	3979.0 (21.1)	3626.4 (21.2)	352.5 (20.9)		2245.0 (25.9)	1840.5 (25.5)	404.5 (28.0)	
all year	6469.8 (34.4)	5892.5 (34.4)	577.3 (34.2)		3747.3 (43.3)	3046.4 (42.2)	700.9 (48.6)	
Occupation by group (N (%))				0.053				<0.001
agricultural - self employed	7788.0 (41.4)	7129.1 (41.6)	658.9 (39.0)		3169.9 (36.6)	2652.2 (36.8)	517.7 (35.9)	
army	391.7 (2.1)	345.3 (2.0)	46.4 (2.7)		201.9 (2.3)	169.0 (2.3)	32.9 (2.3)	
clerical	65.7 (0.3)	62.2 (0.4)	3.5 (0.2)		93.5 (1.1)	73.4 (1.0)	20.1 (1.4)	
not working	5215.9 (27.7)	4781.9 (27.9)	434.0 (25.7)		1539.5 (17.8)	1408.5 (19.5)	131.1 (9.1)	
professional/technical/managerial	531.1 (2.8)	480.5 (2.8)	50.6 (3.0)		919.9 (10.6)	764.9 (10.6)	155.0 (10.7)	
sales	4236.3 (22.5)	3793.4 (22.1)	442.9 (26.2)		665.7 (7.7)	496.5 (6.9)	169.2 (11.7)	
services	460.1 (2.4)	418.7 (2.4)	41.4 (2.5)		1021.0 (11.8)	828.4 (11.5)	192.7 (13.3)	
unskilled manual	66.8 (0.4)	57.7 (0.3)	9.1 (0.5)		211.2 (2.4)	170.7 (2.4)	40.5 (2.8)	

Table 10. Full logit model of mobility among men including the variables from both the final demographic model and the final employment model.

Variable	Estimate	P-Value	Odds ratio (95% CI)
Intercept	-2.751897	0.0000	0.06 (0.04, 0.11)
Employment sector			
no work			
agriculture	0.5470711	0.0000	1.73 (1.38, 2.17)
sales	1.2281688	0.0000	3.41 (2.61, 4.48)
professional/technical/managerial	0.5413067	0.0003	1.72 (1.28, 2.30)
army	0.8873613	0.0001	2.43 (1.55, 3.71)
services	0.8960443	0.0000	2.45 (1.90, 3.17)
manual	1.050962	0.0000	2.86 (2.13, 3.84)
clerical	1.1687657	0.0000	3.22 (1.80, 5.53)
others	-9.773728	0.9414	0.00 (NA, 28.70)
Educational Attainment			
no education			
incomplete primary	-0.173124	0.3077	0.84 (0.61, 1.18)
complete primary	-0.015988	0.9337	0.98 (0.68, 1.44)
incomplete secondary	0.3843602	0.0137	1.47 (1.09, 2.01)
complete secondary	0.3819653	0.0242	1.47 (1.06, 2.06)
higher	0.6240364	0.0013	1.87 (1.28, 2.75)
Rurality			
urban			
small city	0.4138832	0.0420	1.51 (1.00, 2.23)
town	0.6990154	0.0000	2.01 (1.52, 2.67)
countryside	0.953782	0.0000	2.60 (1.97, 3.44)
Wealth Index			
poorest			
poorer	0.2419661	0.0136	1.27 (1.05, 1.54)
middle	0.1555164	0.1130	1.17 (0.96, 1.42)
richer	0.2655656	0.0113	1.30 (1.06, 1.60)
richest	0.2276838	0.1540	1.26 (0.92, 1.72)
Marital status			
married			
never in union	-0.589652	0.0000	0.55 (0.46, 0.67)
living with partner	-0.186022	0.0481	0.83 (0.69, 1.00)
separated	0.1245726	0.4814	1.13 (0.79, 1.59)
widowed	0.5353541	0.0350	1.71 (1.02, 2.77)
divorced	-0.229946	0.5319	0.79 (0.36, 1.55)
Age (years)	-0.014652	0.0000	0.99 (0.98, 0.99)

Table 11. Relative fit of logit models of mobility among men. Models were evaluated using the AIC statistic. Number of parameters refers to the full model; many categorical variables contain more than two levels and are therefore parameterized using 2+ terms.

Model	Number of parameters	AIC
Final demographic model	19	7248
Employment Status (full time, seasonal, occasional, no work)	4	7343
Employment Sector (9 possible categories including 8 sectors + no work)	9	7324
Employment Status *Employment Sector	24	7327
Final demographic + employment model	27	7162

Table 12. Demographic characteristics of all adults in the 2012 Enquéte 1-2-3 survey, stratified by migrants and non-migrants.

	Missing	Migrant	Non-migrant	Overall
N	*859*	*15448*	*43395*	*59702*
Household size (mean (SD))	6.27 (3.18)	6.24 (3.08)	6.25 (3.04)	6.24 (3.05)
Age (mean (SD)) **	36.65 (16.00)	37.84 (15.84)	33.26 (14.90)	34.49 (15.30)
Rurality of current res. (%) **				
City	211 (24.6)	3044 (19.7)	7508 (17.3)	10763 (18.0)
Village	328 (38.2)	4711 (30.5)	13872 (32.0)	18911 (31.7)
Rural	320 (37.3)	7693 (49.8)	22015 (50.7)	30028 (50.3)
Educational attainment (%) **				
None	149 (17.3)	3145 (20.4)	9431 (21.7)	12725 (21.3)
Primary	186 (21.7)	4055 (26.2)	10727 (24.7)	14968 (25.1)
Secondary	399 (46.4)	6826 (44.2)	19812 (45.7)	27037 (45.3)
Non-formal Program	6 (0.7)	52 (0.3)	124 (0.3)	182 (0.3)
University	90 (10.5)	1029 (6.7)	2282 (5.3)	3401 (5.7)
Post-University	2 (0.2)	21 (0.1)	64 (0.1)	87 (0.1)
Professional (INPP)	1 (0.1)	104 (0.7)	208 (0.5)	313 (0.5)
Other	0 (0.0)	34 (0.2)	70 (0.2)	104 (0.2)
Missing	26 (3.0)	182 (1.2)	677 (1.6)	885 (1.5)
Marital status (%) **				
Married- monogomous	420 (48.9)	8523 (55.2)	19786 (45.6)	28729 (48.1)
Single, never married	249 (29.0)	3297 (21.3)	15244 (35.1)	18790 (31.5)
Divorced	48 (5.6)	629 (4.1)	1697 (3.9)	2374 (4.0)
Married- polygamous	45 (5.2)	1081 (7.0)	2195 (5.1)	3321 (5.6)
Common law	33 (3.8)	895 (5.8)	2218 (5.1)	3146 (5.3)
Widowed	38 (4.4)	1006 (6.5)	2174 (5.0)	3218 (5.4)
Missing	26 (3.0)	17 (0.1)	81 (0.2)	124 (0.2)

Figure 21. IDSR case definition for yellow fever from the technical manual.

Yellow fever

Standard case definition

Suspected case:
Any person with acute onset of fever, with jaundice appearing within 14 days of onset of the first symptoms.

Probable case:
A suspected case

AND

One of the following
- Epidemiological link to a confirmed case or an outbreak
- Positive post-mortem liver histopathology

Confirmed case:
A probable case

AND

One of the following
- Detection of YF-**specific*** IgM
- Detection of four-fold increase in YF IgM and/or IgG antibody titres between acute and convalescent serum samples
- Detection of YFV-**specific*** neutralizing antibodies

**YF-specific means that antibody tests (such as IgM or neutralizing antibody) for other prevalent flavivirus are negative. This testing should include at least IgM for Dengue and West Nile and may include other flavivirus depending on local epidemiology.*

OR

One of the following
- Detection of YF virus genome in blood or other organs by PCR
- Detection of yellow fever antigen in blood, liver or other organs by immunoassays
- Isolation of the yellow fever virus

Figure 22. Comparison of IDSR reports and cases between YFV and cholera in the DRC. The left column contains maps relating to YFV, and the right column contains maps relating to cholera. The top row shows the number of years with at least one IDSR report filed (positive or negative) for the given disease, for each health zone. The bottom row shows the number of years with at least one case of the given disease, for each health zone.

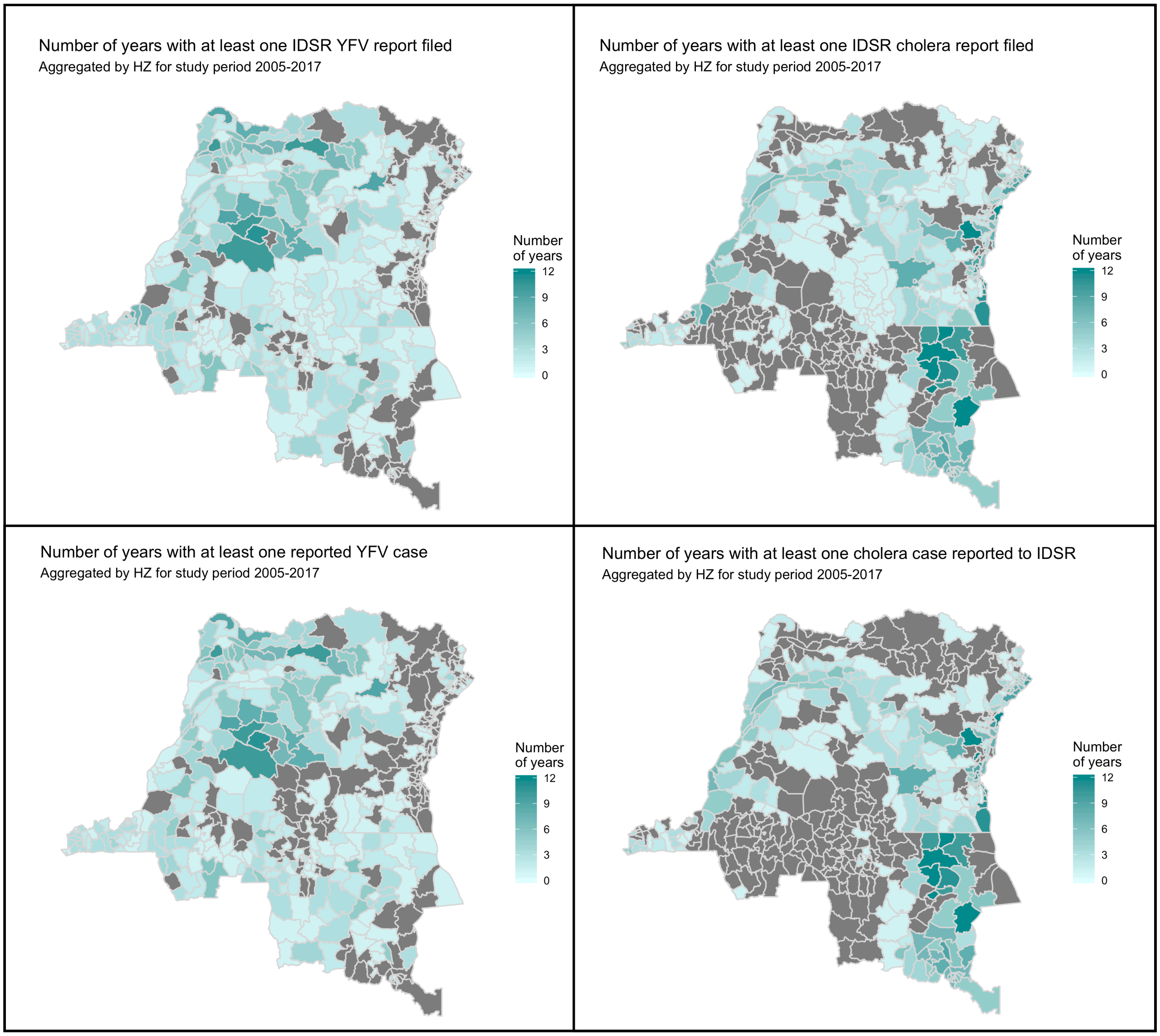
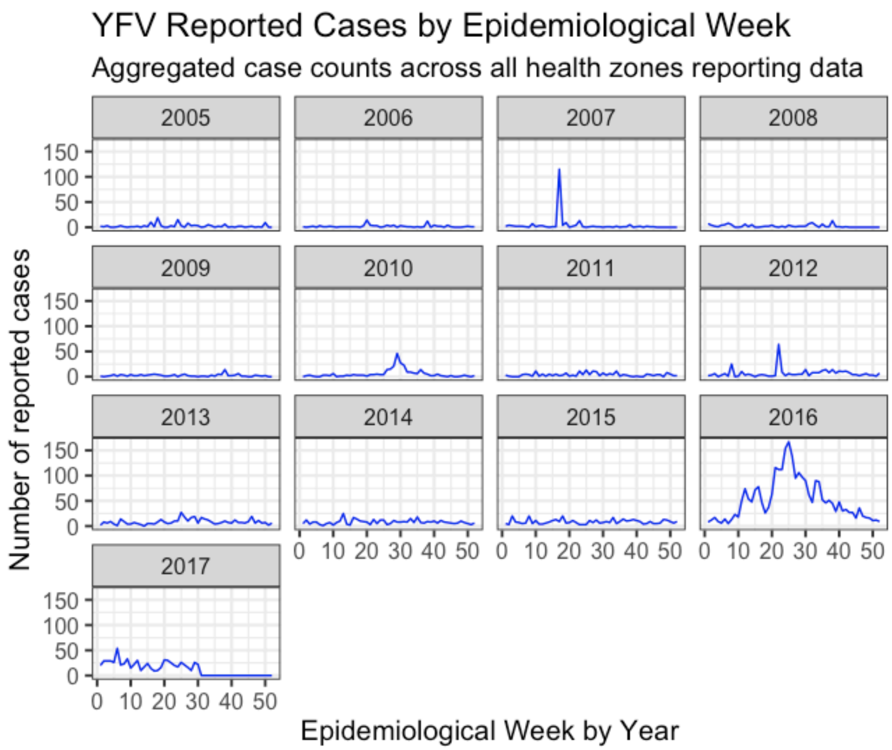


Figure 23. Cases and deaths due to suspected YFV in the DRC, 2005-2017.



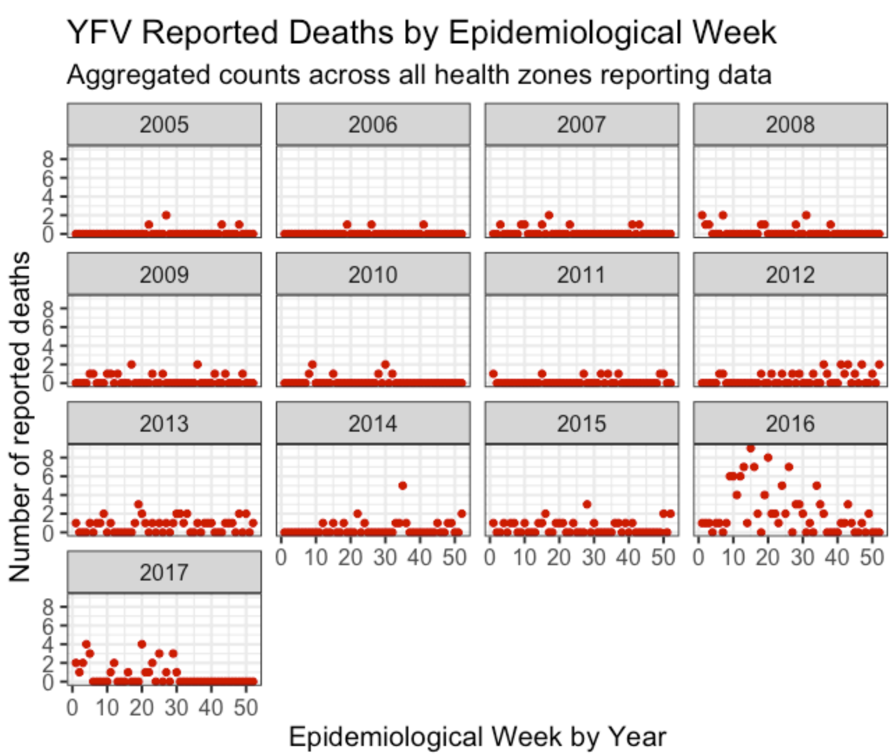
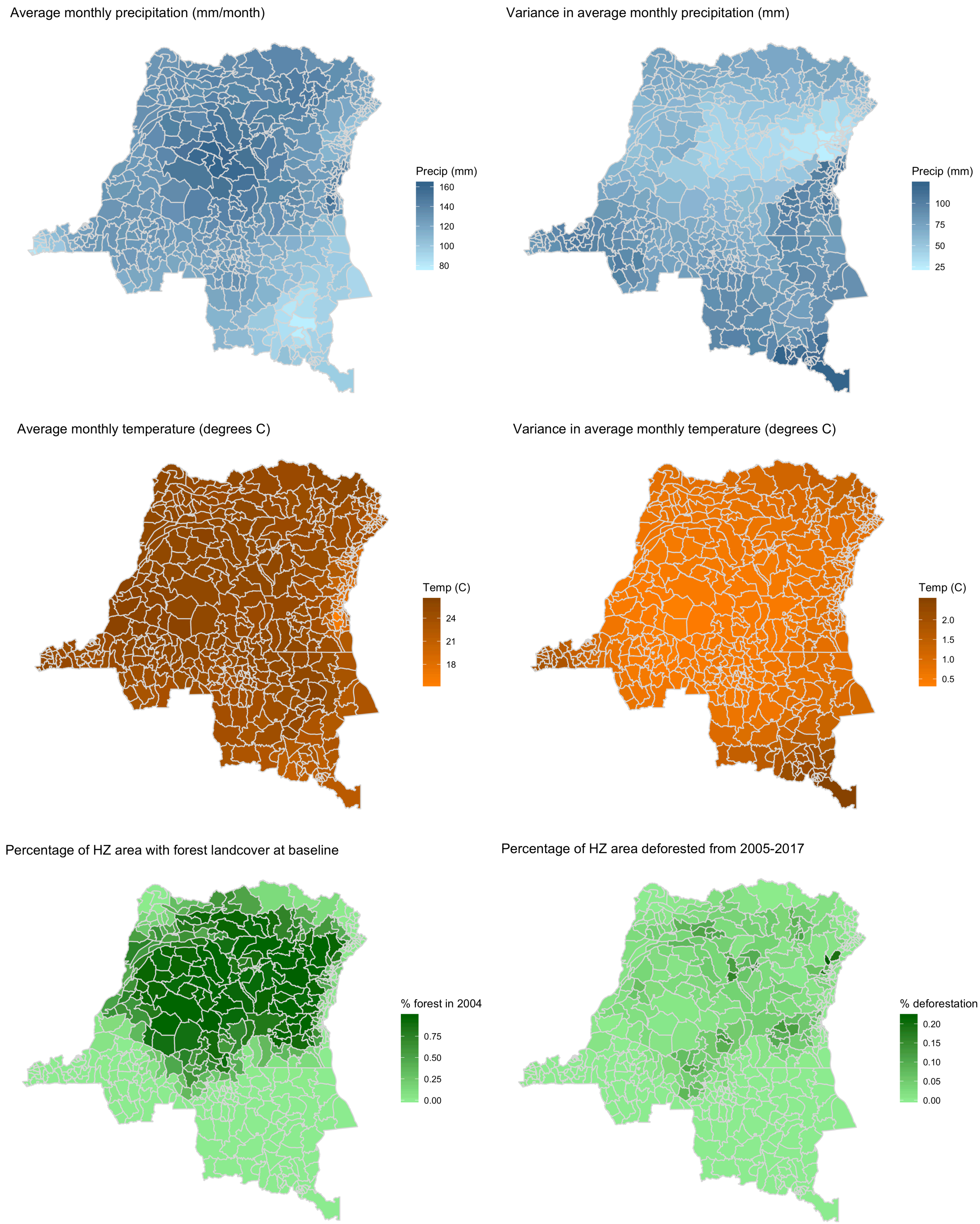


Figure 24. Environmental correlates included in YFV regression models.



REFERENCES

Abdi, H. (2010). Coefficient of Variation. In N. Salkind (Ed.), *Encyclopedia of Research Design*. SAGE Publications, Inc. https://doi.org/10.4135/9781412961288.n56

Ahmed, Q. A., & Memish, Z. A. (2017). Yellow fever from Angola and Congo: A storm gathers. *Tropical Doctor*, *47*(2), 92–96. https://doi.org/10.1177/0049475517699726

Ahuka-Mundeke, S., Casey, R. M., Harris, J. B., Dixon, M. G., Nsele, P. M., Kizito, G. M., Umutesi, G., Laven, J., Paluku, G., Gueye, A. S., Hyde, T. B., Sheria, G. K. M., Muyembe-Tanfum, J.-J., & Staples, J. E. (2018). Immunogenicity of Fractional-Dose Vaccine during a Yellow Fever Outbreak—Preliminary Report. *New England Journal of Medicine*, NEJMoa1710430. https://doi.org/10.1056/NEJMoa1710430

Andre, M. (2017). Landscape ecological consequences of the (sub)urbanization process in an African city: Lubumbashi (Democratic Republic of Congo). In *Doctoral Dissertation*.

Angula, N., & Dlodlo, N. (2018). Enabling Semantic Interoperability of Disease Surveillance Data in Health Information Exchange Systems for Community Health Workers. In H. Florez, C. Diaz, & J. Chavarriaga (Eds.), *Applied Informatics* (pp. 119–130). Springer International Publishing. https://doi.org/10.1007/978-3-030-01535-0_9

Balcan, D., Colizza, V., Gonçalves, B., Hu, H., Ramasco, J. J., & Vespignani, A. (2009). Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences*, *106*(51), 21484 LP – 21489.

Barrett, A. D. T. (2017). Yellow fever live attenuated vaccine: A very successful live attenuated vaccine but still we have problems controlling the disease. *Vaccine*, *35*(44), 5951–5955. https://doi.org/10.1016/j.vaccine.2017.03.032

Barrett, A. D. T. (2018). The reemergence of yellow fever. *Science*, *361*(6405), 847–848. https://doi.org/10.1126/science.aau8225

Bell, M. (2000). Comparing temporary mobility with permanent migration. *Tourism Geographies*, 87–107.

Bhaduri, B., Bright, E., Coleman, P., & Urban, M. L. (2007). LandScan USA: A high-resolution geospatial and temporal modeling approach for population distribution and dynamics. *GeoJournal*, *69*(1/2), 103–117. JSTOR.

Bharti, N., Tatem, A. J., Ferrari, M. J., Grais, R. F., Djibo, A., & Grenfell, B. T. (2011). Explaining Seasonal Fluctuations of Measles in Niger Using Nighttime Lights Imagery. *Science*, *334*(6061), 1424 LP – 1427.

Brockmann, D., & Helbing, D. (2013). The Hidden Geometry of Complex, Network-Driven Contagion Phenomena. *Science*, *342*(6164), 1337 LP – 1342.

Buckee, C. O., Tatem, A. J., & Metcalf, C. J. E. (2017). Seasonal Population Movements and the Surveillance and Control of Infectious Diseases. *Trends in Parasitology*, *33*(1), 10–20. https://doi.org/10.1016/j.pt.2016.10.006

Carrington, C. V. F., & Auguste, A. J. (2013). Evolutionary and ecological factors underlying the tempo and distribution of yellow fever virus activity. *Infection, Genetics and Evolution*, *13*, 198–210. https://doi.org/10.1016/J.MEEGID.2012.08.015

Casey, R. M., Harris, J. B., Ahuka-Mundeke, S., Dixon, M. G., Kizito, G. M., Nsele, P. M., Umutesi, G., Laven, J., Kosoy, O., Paluku, G., Gueye, A. S., Hyde, T. B., Ewetola, R., Sheria, G. K. M., Muyembe-Tamfum, J.-J., & Staples, J. E. (2019). Immunogenicity of Fractional-Dose Vaccine during a Yellow Fever Outbreak—Final Report. *New England Journal of Medicine*, *381*(5), 444–454. https://doi.org/10.1056/NEJMoa1710430

CDC. (2019). *MMWR Week Overview*. MMWR Weeks Fact Sheet and Calendar.

CIA. (2019). *Africa: Congo, Democratic Republic of the—The World Factbook—Central Intelligence Agency*. https://www.cia.gov/library/publications/the-world-factbook/geos/cg.html

Dery, D. B., Brown, C., Asante, K. P., Adams, M., Dosoo, D., Amenga-Etego, S., Wilson, M., Chandramohan, D., Greenwood, B., & Owusu-Agyei, S. (2010). Patterns and seasonality of malaria transmission in the forest-savannah transitional zones of Ghana. *Malaria Journal*, *9*(1), 314. https://doi.org/10.1186/1475-2875-9-314

Emch, M., Root, E. D., & Carrel, M. (2017). *Health and Medical Geography, Fourth Edition*. Guilford Publications.

Eric, M. M. N., Shouyu, C., & Qin, Z. li. (2010). Journal of sustainable development JSD. In *Journal of Sustainable Development* (Vol. 3, Issue 2). CCSE.

Faria, N. R., Kraemer, M. U. G., Hill, S. C., Jesus, J. G. de, Aguiar, R. S., Iani, F. C. M., Xavier, J., Quick, J., Plessis, L. du, Dellicour, S., Thézé, J., Carvalho, R. D. O., Baele, G., Wu, C.-H., Silveira, P. P., Arruda, M. B., Pereira, M. A., Pereira, G. C., Lourenço, J., … Pybus, O. G. (2018). Genomic and epidemiological monitoring of yellow fever virus transmission potential. *Science*, *361*(6405), 894–899. https://doi.org/10.1126/science.aat7115

Faria, Nuno R., Vidal, N., Lourenco, J., Raghwani, J., Sigaloff, K. C. E., Tatem, A. J., Vijver, D. A. M. van de, Pineda-Peña, A.-C., Rose, R., Wallis, C. L., Ahuka-Mundeke, S., Muyembe-Tamfum, J.-J., Muwonga, J., Suchard, M. A., Wit, T. F. R. de, Hamers, R. L., Ndembi, N., Baele, G., Peeters, M., … Dellicour, S. (2019). Distinct rates and patterns of spread of the major HIV-1 subtypes in Central and East Africa. *PLOS Pathogens*, *15*(12), e1007976. https://doi.org/10.1371/journal.ppat.1007976

Findlay, A. M. (2011). Migrant destinations in an era of environmental change. *Global Environmental Change*, *21*, S50–S58. https://doi.org/10.1016/J.GLOENVCHA.2011.09.004

Garcia, A. J., Pindolia, D. K., Lopiano, K. K., & Tatem, A. J. (2015). Modeling internal migration flows in sub-Saharan Africa using census microdata. *Migration Studies*, *3*(1), 89–110.

Garrett, N., & Lintzer, M. (2010). Can Katanga's mining sector drive growth and development in the DRC? *Journal of Eastern African Studies*, *4*(3), 400–424. https://doi.org/10.1080/17531055.2010.517408

Garske, T., Van Kerkhove, M. D., Yactayo, S., Ronveaux, O., Lewis, R. F., Staples, J. E., Perea, W., Ferguson, N. M., & Committee, for the Y. F. E. (2014). Yellow Fever in Africa: Estimating the Burden of Disease and Impact of Mass Vaccination from Outbreak and Serological Data. *PLOS Medicine*, *11*(5), e1001638.

Geenen, S. (2014). Dispossession, displacement and resistance: Artisanal miners in a gold concession in South-Kivu, Democratic Republic of Congo. *Resources Policy*, *40*, 90–99. https://doi.org/10.1016/j.resourpol.2013.03.004

Gostin, L. O., DeBartolo, M. C., & Friedman, E. A. (2016). The International Health Regulations 10 years on: The governing framework for global health security. *The Lancet*, *386*(10009), 2222–2226. https://doi.org/10.1016/S0140-6736(15)00948-4

Hamlet, A., Jean, K., Perea, W., Yactayo, S., Biey, J., Van Kerkhove, M., Ferguson, N., & Garske, T. (2018). The seasonal influence of climate and environment on yellow fever transmission across Africa. *PLOS Neglected Tropical Diseases*, *12*(3), e0006284. https://doi.org/10.1371/journal.pntd.0006284

Hansen, M.C., A. Krylov, A. Tyukavina, P.V. Potapov, S. Turubanova, B. Zutta, S. Ifo, B. Margono, F. Stolle, and R. Moore. (2016. Humid tropical forest disturbance alerts using Landsat data. Environmental Research Letters, 11 (3). Accessed through Global Forest Watch on 06/03/2020. www.globalforestwatch.org

Harris, I., Jones, P. D., Osborn, T. J., & Lister, D. H. (2014). Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset. *International Journal of Climatology*, *34*(3), 623–642. https://doi.org/10.1002/joc.3711

Hay, S. I., George, D. B., Moyes, C. L., & Brownstein, J. S. (2013). Big data opportunities for global infectious disease surveillance. *PLoS Medicine*, *10*(4), e1001413. https://doi.org/10.1371/journal.pmed.1001413

Hoff, N. A., Doshi, R. H., Colwell, B., Kebela-Illunga, B., Mukadi, P., Mossoko, M., Spencer, D., Muyembe-Tamfum, J.-J., Okitolonda-Wemakoy, E., Lloyd-Smith, J., & Rimoin, A. W. (2017). Evolution of a Disease Surveillance System: An Increase in Reporting of Human Monkeypox Disease in the Democratic Republic of the Congo, 2001–2013. *International Journal of Tropical Disease & Health*, *25*(2). https://doi.org/10.9734/IJTDH/2017/35885

Hristoski, I., & Sotiroski, K. (2012). Conceptual Data Modelling of Modern Human Migration. *Management Information Systems*, *7*(4), 3–12.

Hunter, J. M., & Young, J. C. (1971). Diffusion of influenza in England and Wales. *Annals of the Association of American Geographers*, *61*(4), 637–653.

Kabamba, P. (2010). 'Heart of Darkness': Current images of the DRC and their theoretical underpinning. *Anthropological Theory*, *10*(3), 265–301. https://doi.org/10.1177/1463499610372248

Kelly, J. T. D. (2014). "This mine has become our farmland": Critical perspectives on the coevolution of artisanal mining and conflict in the Democratic Republic of the Congo. *Resources Policy*, *40*, 100–108. https://doi.org/10.1016/j.resourpol.2013.12.003

Kraemer, M. U. G., Faria, N. R., Reiner, R. C., Golding, N., Nikolay, B., Stasse, S., Johansson, M. A., Salje, H., Faye, O., Wint, G. R. W., Niedrig, M., Shearer, F. M., Hill, S. C., Thompson, R. N., Bisanzio, D., Taveira, N., Nax, H. H., Pradelski, B. S. R., Nsoesie, E. O., … Cauchemez, S. (2017a). Spread of yellow fever virus outbreak in Angola and the Democratic Republic of the Congo 2015–16: A modelling study. *The Lancet Infectious Diseases*, *17*(3), 330–338. https://doi.org/10.1016/S1473-3099(16)30513-8

Kraemer, M. U. G., Faria, N. R., Reiner, R. C., Golding, N., Nikolay, B., Stasse, S., Johansson, M. A., Salje, H., Faye, O., Wint, G. R. W., Niedrig, M., Shearer, F. M., Hill, S. C., Thompson, R. N., Bisanzio, D., Taveira, N., Nax, H. H., Pradelski, B. S. R., Nsoesie, E. O., … Cauchemez, S. (2017b). Spread of yellow fever virus outbreak in Angola and the Democratic Republic of the Congo 2015–16: A modelling study. *The Lancet Infectious Diseases*, *17*(3), 330–338. https://doi.org/10.1016/S1473-3099(16)30513-8

Lynch, C., & Roper, C. (2011). The Transit Phase of Migration: Circulation of Malaria and Its Multidrug-Resistant Forms in Africa. *PLOS Medicine*, *8*(5), e1001040.

Mabaso, M. L. H., Craig, M., Vounatsou, P., & Smith, T. (2005). Towards empirical description of malaria seasonality in southern Africa: The example of Zimbabwe. *Tropical Medicine & International Health*, *10*(9), 909–918. https://doi.org/10.1111/j.1365-3156.2005.01462.x

Maclin, B. J., Kelly, J. T. D., Perks, R., Vinck, P., & Pham, P. (2017). *Moving to the mines: Motivations of men and women for migration to artisanal and small-scale mining sites in Eastern Democratic Republic of the Congo*. https://doi.org/10.1016/j.resourpol.2016.12.003

Makiala-Mandanda, S., Ahuka-Mundeke, S., Abbate, J. L., Pukuta-Simbu, E., Nsio-Mbeta, J., Berthet, N., Leroy, E. M., Becquart, P., & Muyembe-Tamfum, J.-J. (2018). Identification of Dengue and Chikungunya Cases Among Suspected Cases of Yellow Fever in the Democratic Republic of the Congo. *Vector-Borne and Zoonotic Diseases*, *18*(7), 364–370. https://doi.org/10.1089/vbz.2017.2176

Makiala-Mandanda, S., Le Gal, F., Ngwaka-Matsung, N., Ahuka-Mundeke, S., Onanga, R., Bivigou-Mboumba, B., Pukuta-Simbu, E., Gerber, A., Abbate, J. L., Mwamba, D., Berthet, N., Leroy, E. M., Muyembe-Tamfum, J.-J., & Becquart, P. (2017). High Prevalence and Diversity of Hepatitis Viruses in Suspected Cases of Yellow Fever in the

Democratic Republic of Congo. *Journal of Clinical Microbiology*, *55*(5), 1299–1312. https://doi.org/10.1128/JCM.01847-16

Marivoet, W., & De Herdt, T. (2017). From Figures to Facts: Making sense of socio-economic surveys in the Democratic Republic of the Congo (DRC). *Analysis and Policy Brief N23*, 6.

Marivoet, W., & De Herdt, T. (2018). Tracing Down Real Socio-Economic Trends From Household Data With Erratic Sampling Frames: The Case of the Democratic Republic of the Congo. *Journal of Asian and African Studies*, *53*(4), 532–552. https://doi.org/10.1177/0021909617698842

Marivoet, W., & Herdt, T. D. (2014). Reliable, challenging or misleading? A qualitative account of the most recent national surveys and country statistics in the DRC. *Canadian Journal of Development Studies / Revue Canadienne d'études Du Développement*, *35*(1), 97–119. https://doi.org/10.1080/02255189.2014.877381

May, J. M. (1959). The Ecology of Human Disease. *The Ecology of Human Disease.* https://www.cabdirect.org/cabdirect/abstract/19592703566

Mayer, J. D. (2009). Medical Geography. In T. Brown, S. McLafferty, & G. Moon (Eds.), *A Companion to Health and Medical Geography*. John Wiley & Sons, Incorporated. http://ebookcentral.proquest.com/lib/unc/detail.action?docID=470148

Meade, M. S. (1977). Medical Geography as Human Ecology: The Dimension of Population Movement. *Geographical Review*, *67*(4), 379–393. JSTOR. https://doi.org/10.2307/213623

Monath, T. P., & Vasconcelos, P. F. C. (2015). Yellow fever. *Journal of Clinical Virology*, *64*, 160–173. https://doi.org/10.1016/j.jcv.2014.08.030

Otshudiema, J. O., Ndakala, N. G., Mawanda, E. K., Tshapenda, G. P., Kimfuta, J. M., Nsibu, L.-R. N., Gueye, A. S., Dee, J., Philen, R. M., Giese, C., Murrill, C. S., Arthur, R. R., & Kebela, B. I. (2017). Yellow Fever Outbreak—Kongo Central Province, Democratic Republic of the Congo, August 2016. *MMWR. Morbidity and Mortality Weekly Report*, *66*(12), 335–338. https://doi.org/10.15585/mmwr.mm6612a5

Paules, C. I., & Fauci, A. S. (2017). Yellow Fever—Once Again on the Radar Screen in the Americas. *New England Journal of Medicine*, *376*(15), 1397–1399. https://doi.org/10.1056/NEJMp1702172

Petersen, W. (1958). A General Typology of Migration. *American Sociological Review*, *23*(3), 256–266. https://doi.org/10.2307/2089239

Pindolia, D. K., Garcia, A. J., Wesolowski, A., Smith, D. L., Buckee, C. O., Noor, A. M., Snow, R. W., & Tatem, A. J. (2012). Human movement data for malaria control and elimination strategic planning. *Malaria Journal*, *11*(1), 205. https://doi.org/10.1186/1475-2875-11-205

Prothero, R. M. (1977). Disease and Mobility: A Neglected Factor in Epidemiology*. *International Journal of Epidemiology*, *6*(3), 259–267.

Rebaudet, S., Sudre, B., Faucher, B., & Piarroux, R. (2013). Environmental Determinants of Cholera Outbreaks in Inland Africa: A Systematic Review of Main Transmission Foci and Propagation Routes. *The Journal of Infectious Diseases*, *208*(suppl_1), S46–S54. https://doi.org/10.1093/infdis/jit195

Ruktanonchai, N. W., DeLeenheer, P., Tatem, A. J., Alegana, V. A., Caughlin, T. T., zu Erbach-Schoenberg, E., Lourenço, C., Ruktanonchai, C. W., & Smith, D. L. (2016). Identifying Malaria Transmission Foci for Elimination Using Human Mobility Data. *PLOS Computational Biology*, *12*(4), e1004846.

Rustad, S. A., Østby, G., & Nordås, R. (2016). *Artisanal mining, conflict, and sexual violence in Eastern DRC*. https://doi.org/10.1016/j.exis.2016.01.010

Sanguma, W. (2015). Conflict and Its Effect on Urbanization. *Peace Review*, *27*(4), 507–514. https://doi.org/10.1080/10402659.2015.1094345

Shearer, F. M., Longbottom, J., Browne, A. J., Pigott, D. M., Brady, O. J., Kraemer, M. U. G., Marinho, F., Yactayo, S., de Araújo, V. E. M., da Nóbrega, A. A., Fullman, N., Ray, S. E., Mosser, J. F., Stanaway, J. D., Lim, S. S., Reiner Jr, R. C., Moyes, C. L., Hay, S. I., & Golding, N. (2018). Existing and potential infection risk zones of yellow fever worldwide: A modelling analysis. *The Lancet Global Health*, *6*(3), e270–e278. https://doi.org/10.1016/S2214-109X(18)30024-X

Shearer, F. M., Moyes, C. L., Pigott, D. M., Brady, O. J., Marinho, F., Deshpande, A., Longbottom, J., Browne, A. J., Kraemer, M. U. G., O'Reilly, K. M., Hombach, J., Yactayo, S., de Araújo, V. E. M., da Nóbrega, A. A., Mosser, J. F., Stanaway, J. D., Lim, S. S., Hay, S. I., Golding, N., & Reiner Jr, R. C. (2018a). Global yellow fever vaccination coverage from 1970 to 2016: An adjusted retrospective analysis. *The Lancet Infectious Diseases*, *17*(11), 1209–1217. https://doi.org/10.1016/S1473-3099(17)30419-X

Shearer, F. M., Moyes, C. L., Pigott, D. M., Brady, O. J., Marinho, F., Deshpande, A., Longbottom, J., Browne, A. J., Kraemer, M. U. G., O'Reilly, K. M., Hombach, J., Yactayo, S., de Araújo, V. E. M., da Nóbrega, A. A., Mosser, J. F., Stanaway, J. D., Lim, S. S., Hay, S. I., Golding, N., & Reiner Jr, R. C. (2018b). Global yellow fever vaccination coverage from 1970 to 2016: An adjusted retrospective analysis. *The Lancet Infectious Diseases*, *17*(11), 1209–1217. https://doi.org/10.1016/S1473-3099(17)30419-X

Stoddard, S. T., Morrison, A. C., Vazquez-Prokopec, G. M., Paz Soldan, V., Kochel, T. J., Kitron, U., Elder, J. P., & Scott, T. W. (2009). The Role of Human Movement in the Transmission of Vector-Borne Pathogens. *PLoS Neglected Tropical Diseases*, *3*(7), e481. https://doi.org/10.1371/journal.pntd.0000481

Tamm, H. (2016). The Origins of Transnational Alliances: Rulers, Rebels, and Political Survival in the Congo Wars. *International Security*, *41*(1), 147–181. https://doi.org/10.1162/ISEC_a_00252

Tatem, A. J. (2014). Mapping population and pathogen movements. *International Health*, *6*(1), 5–11.

Tatem, A. J. (2017). WorldPop, open data for spatial demography. *Scientific Data*, *4*, 170004.

Tsurukawa, N., Prakash, S., & Manhart, A. (2011). *Social impacts of artisanal cobalt mining in Katanga, Democratic Republic of Congo*.

UN OCHA. (2018a). *Democratic Republic of Congo: Internally Displaced Persons and Returnees as of 31 December 2017*.

UN OCHA. (2018b). *Democratic Republic of Congo: Internally Displaced Persons and Returnees as of 31 December 2017*.

UNHCR, U. N. H. C. for. (2014). *UNHCR Global Trends 2014*. UNHCR. https://www.unhcr.org/statistics/country/556725e69/unhcr-global-trends-2014.html

United Nations. (2016). *United Nations Statistics Division—Environment Statistics*. https://unstats.un.org/unsd/environment/totalarea.htm

Vasconcelos, P. F. C., & Monath, T. P. (2016). Yellow Fever Remains a Potential Threat to Public Health. *Vector-Borne and Zoonotic Diseases*, *16*(8), 566–567. https://doi.org/10.1089/vbz.2016.2031

Vereinte Nationen, & United Nations Population and Housing Programme (Eds.). (2017). *Principles and recommendations for population and housing censuses: 2020 round* (Revision 3). United Nations.

Vlassenroot, K., & Huggins, C. (2004a). Land, Migration and Conflict in Eastern D.R. Congo. *Eco-Conflicts*, *3*(4), 1–4.

Vlassenroot, K., & Huggins, C. (2004b). Land, Migration and Conflict in Eastern D.R. Congo. *Eco-Conflicts*, *3*(4), 1–4.

Warner, K., Hamza, M., Oliver-Smith, A., Renaud, F., & Julca, A. (2010). Climate change, environmental degradation and migration. *Natural Hazards*, *55*(3), 689–715. https://doi.org/10.1007/s11069-009-9419-7

Wasserman, S., Tambyah, P. A., & Lim, P. L. (2016). Yellow fever cases in Asia: Primed for an epidemic. *International Journal of Infectious Diseases*, *48*, 98–103. https://doi.org/10.1016/j.ijid.2016.04.025

Wesolowski, A., Buckee, C. O., Pindolia, D. K., Eagle, N., Smith, D. L., Garcia, A. J., & Tatem, A. J. (2013). The Use of Census Migration Data to Approximate Human Movement Patterns across Temporal Scales. *PLOS ONE*, *8*(1), e52971.

WHO. (2015, December 15). *Cholera – Democratic Republic of the Congo*. World Health Organization Disease Outbreak News. http://www.who.int/csr/don/15-december-2015-cholera-drc/en/

WHO. (2017a). *The EYE Strategy addresses three important global health agendas*. WHO. http://www.who.int/csr/disease/yellowfev/eye-strategy-and-global-agendas/en/

WHO. (2017b). *Yellow fever – Brazil*. World Health Organization Disease Outbreak News. http://www.who.int/csr/don/24-november-2017-yellow-fever-brazil/en/

WHO. (2018). *Yellow Fever*. WHO | Regional Office for Africa. https://www.afro.who.int/health-topics/yellow-fever

Wilder-Smith, A., & Leong, W. Y. (2017). Importation of yellow fever into China: Assessing travel patterns. *Journal of Travel Medicine*, *24*(4). https://doi.org/10.1093/jtm/tax008

Woodall, J. P., & Yuill, T. M. (2016). Why is the yellow fever outbreak in Angola a 'threat to the entire world'? *International Journal of Infectious Diseases*, *48*, 96–97. https://doi.org/10.1016/j.ijid.2016.05.001

World Bank. (2018). *Democratic Republic of Congo Urbanization Review*. World Bank.

World Health Organization, & Centers for Disease Control. (2010). *Technical Guidelines for Integrated Disease Surveillance and Response in the African Region*. https://stacks.cdc.gov/view/cdc/11943

Wu, J. T., Peak, C. M., Leung, G. M., & Lipsitch, M. (2016). Fractional dosing of yellow fever vaccine to extend supply: A modelling study. *The Lancet*, *388*(10062), 2904–2911. https://doi.org/10.1016/S0140-6736(16)31838-4

Zhao, S., Stone, L., Gao, D., & He, D. (2018). Modelling the large-scale yellow fever outbreak in Luanda, Angola, and the impact of vaccination. *PLOS Neglected Tropical Diseases*, *12*(1), e0006158. https://doi.org/10.1371/journal.pntd.0006158