Using Cell Phone Data to Improve Disease Targeting and Mitigate the Negative Externality of Internal Population Movement

Sveta Milusheva

March 21, 2016

Abstract

Population mobility within a country can lead to a number of externalities, from the spread of information and goods, to the transmission of disease. Yet until recently, the only data that would allow the study of population movement within a country for large parts of the population were census data or other survey data which only provide a snapshot of movement and do not make it possible to study short term movements and their consequences at a high frequency. Using a unique dataset of Call Detail Records (CDRs) within Senegal for 9 million users over the course of a year, this paper is able to directly study the relationship between mobility and the spread of disease. In addition, using Senegal as a case study, the paper demonstrates how this relationship can be used to specifically target certain areas of the country which are the leading importers of the disease to low malaria areas of the country. This paper also contributes to the literature on movement and disease more generally, demonstrating how as areas approach elimination of a disease, environmental factors decrease in importance and human factors like travel play an increasing role in the propagation of the disease. This highlights the importance of differentiated interventions based on the elimination stage of the area in order to achieve the most cost effective drop in the disease burden.

1 Introduction

Population mobility within a country can lead to a number of externalities, from the spread of information and goods, to the transmission of disease. While the link between travel and disease has been studied in the epidemiological literature at the level of a single village, or a few villages, few studies have looked at the effect of travel on a larger scale, such as an entire country. The main reason is that data on short term movement is difficult to collect, so that studies that have tried to study the spread of disease due to movement on a larger scale have had to use rough proxies for movement. Yet as transportation infrastructure improves and people become more mobile, it is critical to understand the link between this mobility and disease in order to mitigate the negative externality, reduce the disease burden, and prevent the spread of devastating epidemics.

This study is in a unique position to study the effect of short term movement on disease due to an extensive dataset of cell phone usage for 9 million users in Senegal over the course of 2013. This dataset provides the approximate location of all calls and texts for the users, which makes it possible to trace changes in location over time and to extract patterns of movement between different areas. This is ideal for measuring short term movement on a national scale, which cannot be tracked with most large data sources like the census.

While internal movement can impact all infectious diseases, the paper focuses on malaria. Malaria infected almost 200 million people in 2013 and has negative economic consequences for the households affected as well as for the governments trying to fight the disease (*World Malaria Report* 2014,Shepard et al. 1991, Chuma, Thiede, and Molyneux 2006). In addition, many studies have found movement of infected individuals to be a key factor for the resurgence or outbreak of malaria (Cohen et al. 2012, Lu et al. 2014). This disease is particularly important to study in Senegal, because malaria is the second deadliest disease in the country, and is the focus of an eradication campaign in the north of the country. In the context of studying the relationship between movement and disease, it is an especially apt case because in the north of the country, which is the region that is the focus of this study, movement plays the biggest role in the propagation of the disease. In one district where epidemiological survey data was collected, it was found that since 2012, around 70% of all cases in the district have been imported through travel, and these have contributed to more than half of the secondary cases in the district.

Therefore, in order for the Senegalese Ministry of Health to use the resources it is putting into this eradication campaign in the most effective way, it is necessary to understand the travel patterns into the north of the country, the impact of travel on malaria cases, and the areas of the country that contribute the most to the importation of new cases. Utilizing the cell phone data, I am able to study these three questions and provide analysis that can help contribute to the eradication of the disease. While previous studies have used cell phone data to determine areas that are sources of malaria and areas with higher risk due to movement (Tatem, Qiu, et al. 2009, Wesolowski et al. 2012, Enns and Amuasi 2013), this study goes a step further to quantify the effect of movement and determine the timing of transmission by combining the cell phone data with high frequency data on number of malaria cases. This type of analysis has not been done on a national scale before, but is necessary for the effective and efficient distribution of resources by the government.

Several findings come out of this study. While malaria is a disease that is closely linked to environmental factors, there is a transition that occurs as areas reach a very low level of malaria and the environmental factors no longer play as important a role in the propagation of the disease. Instead, in the very low malaria areas, population movement becomes a more important factor for the spread of disease. In addition, I demonstrate that only certain types of movement are linked with increases in malaria prevalence in the low malaria areas. Only people entering from very high malaria areas impact the prevalence of the area they enter. This information is helpful in starting to think about where policies might be most effective in preventing the spread of malaria to places that are close to eradication. Finally, in taking this a step further, I demonstrate how decreases in malaria prevalence among travelers from different parts of the country, which could be brought about through a number of policies, would impact malaria prevalence in the north of the country in order to determine which areas might be most important to target.

The paper begins with some background on the link between migration and the spread of disease. Section 3 describes the data on malaria and population movement. Section 4 outlines the empirical analysis and section 5 provides the results from the analysis. Section 6 looks at some potential policy implications and the paper concludes with section 7.

2 Movement and Malaria

The link between movement and the spread of communicable diseases is not a new topic (Prothero 1977, Balcan et al. 2009, Huang, Tatem, et al. 2013, Tatem and D. L. Smith 2010). Specifically, a number of case-control studies have been done looking at the link between malaria and travel in a specific location. Researchers select a group of people that is diagnosed with malaria and a comparable group that is not and then conduct a survey that asks about travel history, along with other demographic characteristics that could contribute to malaria contraction (Siri et al. 2010, Osorio, Todd, and Bradley 2004, Yukich et al. 2013). They find that recent travel, from 8-14 days up to 30 days prior, is one of the main risk factors for contracting malaria. These types of case-control studies have only been done on single locations due to the cost and time necessary to collect travel history data. Therefore, they suffer from a lack of external validity and inability to be applied to general models of travel in the country that could be used to inform national policy.

Some studies have used census data or national surveys to measure migration in order to describe migration routes and how these relate to the presence of malaria in different parts of a country (Lynch and Roper 2011, Stoddard et al. 2009). Yet the migration data available, especially for internal migration, is often not sufficiently high resolution to establish a link between internal movement and the spread of a disease. In addition, the movement captured by surveys and the census often times misses short term movements and cyclical migration (Deshingkar and Grimm 2004). In economics, attempts have been made to study the effect of movement on spread of disease through variables used to proxy for movement. Adda 2015 studies the effect of school closings and railway strikes on the spread of the flu, diarrhea and chickenpox in France. Oster 2012 studies the effect of exports on the spread of HIV, with the main mechanism being the travel of truckers. Yet in both of these papers, only one or two types of mobility are potentially measured using proxies and it is difficult to make any conclusions about the effect of mobility more generally.

In particular in Senegal, data on movement is only available nationally through the Senegalese Survey of Households (ESAM) and the Census, both of which only have data on migration and no measure of short term visits. Therefore, the only empirical study of movement and malaria is a project done in Richard Toll, one of the districts in Senegal that is covered in the data used for this paper. It tracked malaria cases over 12 weeks and used a questionnaire to learn more about how malaria was spreading (Littrell et al. 2013). The study found that one of the main risk factors for contracting malaria was travel that entailed an overnight stay. A country-wide study of malaria that includes regions with different levels of prevalence has not been done due to the prohibitive cost of tracing and interviewing all malaria cases.

Le Menach et al. 2011 is the closest paper to this research. Combining cell phone data for Zanzibar as well as a dynamic mathematical model of importation and transmission rates, they study the transmission from residents traveling to malaria endemic regions as well as visitors and immigrants coming from endemic regions. They find that residents traveling to malaria endemic regions contribute 15 times more imported cases than infected visitors. They also estimate the malaria importation rate to be 1.6 incoming infections per 1000 inhabitants per year. This study is limited in that it is focused on an island that has an extremely low rate of malaria, which makes the ability to generalize to areas that might have higher rates of malaria difficult. In addition, it contains no measure of number of malaria cases and is only based on theoretical predictions of the prevalence and reproductive rates, which make a number of assumptions that could affect the estimates. Finally, the paper is attempting to estimate the total number of malaria cases generated over the span of the disease, which is estimated to last for 200 days within a person. In contrast, the goal of this paper is to estimate the number of monthly cases generated by movement and the timing of those cases because that is most relevant for policy makers interested in targeting a specific time period when the largest impact is expected.

3 Data

Malaria Data

The data used to measure malaria prevalence comes from the Programme National de Lutte Contre le Paludisme (PNLP) (*Bulletin de Surveillance Sentinelle du Paludisme No 1-46, 2013* 2013). This national program, which has the goal of controlling malaria in Senegal, has been collecting case data at different spatial and temporal levels. There is monthly data by health district starting in 2012 for the whole country, monthly data at the health post level for eight health districts starting in 2013 and weekly data at the health post level for four health districts starting in 2013. The cases reported are all new cases in the respective month or week. All potential malaria cases are tested using a Rapid Diagnostic Test (RDT); therefore, the data consists of confirmed malaria cases.

This paper utilizes the data for health posts at the monthly level. The reason for this is that it provides a mix of both lower and higher malaria health posts (while the weekly health post data covers mostly higher malaria health posts). It also makes it possible to study the effect of movement at a finer scale than the health district, which is important since different areas within a district experience different movement patterns and different levels of malaria, and this variation would be missed if studied at the health district level. In addition, the districts for which there is monthly data at the healthpost level comprise almost the entirety of the northern part of Senegal, which is the area where the Ministry of Health is aiming to eradicate malaria. Figure 1 shows the health districts for which there is monthly data at the health post level. These health districts are subdivided into areas based on the location of the health posts and cell phone towers. For those healthposts close together (in the same village or neighboring villages) where it would not be possible to distinguish which one individuals in the area would utilize, they were grouped together into one catchment area. The subdivisions are shown on the map in light gray.

Several things should be noted about the data and malaria in Senegal that are important for the analysis. Although several malaria parasites exist worldwide, in Senegal, almost all cases of malaria are due to *P. falciparum* (*World Malaria Report* 2014). The malarial cycle for *P. falciparum* can take several weeks, but importantly, unlike some of the other malarial parasites, it does not have the potential to lie dormant for a period of time, and therefore if symptoms appear, they will happen within a few weeks of the infection and there will be no relapses years later, which happens with some of the other malaria parasites and would significantly complicate the current analysis.[1] In addition, the focus of the analysis on the north of the country further simplifies the modeling of malaria prevalence in this study because immunity is rarely acquired in this part of the country due to the low prevalence. In areas where malaria is endemic, multiple infections with malaria will lead individuals to develop what is known as acquired immunity and means that while a person might still carry the parasite and be able to infect mosquitoes, he or she does not develop symptoms, and therefore would not show up as a case in a health post. In the north of the country since malaria is not endemic but relatively rare, individuals living in this area do not have immunity and do develop symptoms that are usually severe and necessitate the need to seek treatment at a health facility. Furthermore, there are active health workers throughout the region that help to identify cases of malaria even in areas where there is not a health post close by, and these cases are added to

[1] Details on malaria transmission and the biological cycle can be found in Doolan, Dobaño, and Baird 2009, D. L. Smith and McKenzie 2004, Killeen, Ross, and T. Smith 2006, Johnston, D. L. Smith, and Fidock 2013, Wiser 2010).

the case numbers of the health post closest to where they were found. These aspects particular to the region of Senegal studied help to ensure that the healthpost case data is an accurate representation of malaria prevalence in this area.

Cell Phone Records

The data used to measure short term visits come from phone records made available by Sonatel and Orange in the context of the D4D Challenge, a call for projects with the objective to explore the potential of mobile call data to facilitate socio-economic development. The data consist of call and text data for Senegal between January 1, 2013 and December 31, 2013 for all of Sonatel's user base. In 2013, Sonatel had slightly over 9 million unique phone numbers on its network. To put this in perspective, the total population of Senegal in 2013 was around 14 million. The data contains information on all calls and texts made or received by an individual, their time, date and location of the closest cell phone tower, which is what allows for individuals to be tracked in space as they make calls from different tower locations. The data is anonymized, so that while random IDs are provided that make it possible to track the same individual over time, there is no identifying information on the individuals.

Figure 2 demonstrates the total number of calls and texts made each day by the individuals in the dataset. At certain times during the year there are spikes in the number of phone calls. Vertical lines in the graph mark major holidays, which sometimes correspond to the spikes in phone calls. Korité and Tabaski are the two biggest Muslim holidays, known also as Eid al-Fitr and Eid al-Adha respectively, and the graph demonstrates that many people call and text on those holidays. Yet, as the movement graphs will show later, the movement is not directly linked to the number of calls/texts being made, because some of the holidays associated with significant movement are not associated with above average usage patterns.

The current study only focuses on the last phone call/text of the day made or received by individuals in order to determine the locations where individuals reside rather than capturing work locations[2]. First, in order to measure movement a loca-

[2]This method was compared with another method of capturing individual locations used in the literature involving only looking at calls made between 7pm and 7am. The two methods did not differ substantively and the last call of the day method was chosen since it was slightly more inclusive and did not remove individuals or days for individuals when no calls were made in the time frame specified by the alternative method.

tion is assigned to each person for each day. The location of the last call of the day is assigned as the location for that day. In instances where there are days with no calls, the location of the day closest to the one missing where a location is known based on a call is assigned to the missing days. In this way a location is assigned for every day of the year. Movement is defined as a change in location from one health district to another between two consecutive days, with the variable of interest being number of people that enter a district based on this definition of change.

Precipitation Data

There is a large literature that looks at the effect of precipitation on malaria. Therefore, rainfall is included in the model. Data on rainfall come from the NOAA Climate Prediction Center Rainfall Estimation Algorithm Version 2 (RFE 2.0). Data is provided daily with a resolution of 0.1 degree based on a combination of estimates from three different satellites and GTS rain gauge data. For the analysis, the raster data within each health catchment area were averaged on a daily basis to create a daily rainfall measure for the area. A monthly measure for rainfall was created by summing the rainfall for each day in the month. This is due to the fact that previous literature has shown the importance of cumulative rainfall (Silal 2012, Hoshen and Morse 2004).

4 Empirical Analysis

First we look at some descriptive measures of the data. Figure 3 shows annual malaria prevalence throughout Senegal in 2013. We can see that for the most part, the north of the country, which is the focus of this study, has very low malaria and is at the level considered ready for elimination (below 5 cases per 1000). The large heterogeneity in malaria prevalence in the country is ideal for studying the effect of population movement on disease prevalence because it means there are areas of the country that can act as sources for malaria and other areas which are sinks, where new cases arise from individuals that are coming from the high malaria areas.

In Figure 4, the focus is on the eight health districts for which there is monthly data at the healthpost level. The top panel shows cases per month from 2013 to 2015 by district, and we can see that three districts stand out as having a much higher prevalence. These districts are labeled "high malaria" northern districts and this term is included in the model. Panel B focuses on the five low malaria northern districts and we see that all districts, both low and high, have a very seasonal pattern with a single annual peak of malaria occurring around September/October. This seasonality is measured using month and year fixed effects. The seasonality is very closely linked to rainfall though, and in the Appendix this relationship is shown in figure 9. Results are also shown in the appendix where seasonality is explicitly measuring using lagged rainfall rather than fixed effects.

Turning now to look at the movement data, Figure 5 shows the number of people entering a northern healthpost catchment area each day, along with vertical lines marking public holidays and important pilgrimages. We see that the movement patterns largely align with the holidays and pilgrimages, showing that the majority of deviations that we see in movement are arising from special occasions that cause individuals to travel, often to visit family and friends, or else for religious pilgrimages, to visit a sacred site. In Figure 6, this movement is further broken down by the malaria level of the location that the individuals are entering from. Rather than raw numbers, the data is now displayed as deviations from the mean for each group. This breakdown shows that the patterns are actually different for people entering from areas with different levels of malaria. The differences in movement are due to the historical relationships between different areas, and it is important for policy makers to take into account these different patterns since not all movement will lead to high malaria, and so only certain patterns of movement are of interest. From some areas, the movement is much more associated with pilgrimages and holiday seasons. For the individuals coming from high malaria areas, the association seems to be less linked to holidays or pilgrimages. Instead, it is related to the agricultural calendar, since many of the people entering are seasonal workers coming to help with the harvest.While for those coming from low malaria areas, which are the areas that are geographically closest, there does not seem to be much deviation from the variation during the year, instead the flow of movement remains relatively steady throughout the year.

Before modeling the relationship between movement and malaria, first malaria prevalence is modeled based only on seasonality and a few additional variables. This is estimated using the following equation:

$$\text{Cases}_j = exp(\beta_0 + \beta_1 Rain_j + \beta_2 Dist_j + \beta_3 Rain_j * H_j + \beta_4 H_j + \sum_{t=2}^{12} \lambda_t Month_t + \sum_{y=2}^{3} \delta_y Year_y) * Pop_j$$

This model is estimated using a negative binomial regression due to over dispersion in the case data.[3] In addition to monthly variables to capture the seasonality, $(\sum_{t=2}^{12} \lambda_t Month_t)$, I also include year fixed effects $(\sum_{y=2}^{3} \delta_y Year)$. Total rainfall per month is also included. I include a measure for average distance of healthposts in the catchment area to a road, since those healthposts located near a road might receive more patients because it is easier to get to them and helps to act as a measure of treatment seeking behavior and the fact that in some areas higher malaria cases might reflect more people choosing to seek treatment. As already mentioned, in the North of the country a majority of cases should be captures through the community health workers and the fact that cases tend to be more severe and require seeking treatment, but this variable nevertheless is used to capture the possibility that there might be some cases that are not included due to the healthpost being difficult to reach. A variable for whether the healthpost catchment area is in a higher malaria district (H_j), as described earlier, is also included. The high variable is interacted with the rainfall measure to account for the possibility that environmental factors like rainfall might have different effects depending on the malaria level of the area. Population is entered as an offset that allows the model to reflect the rate of cases. Variables are introduced one by one in order to see how they change the model and the fit.

Next, the movement variable from the cell phone data is added. This model can only be estimated for one year of data since cell phone data is only available for 2013.

$$\text{Cases}_j = exp(\alpha_1 \sum_i (Ent_{ji} * Incid_i) + \alpha_2 \sum_i (Ent_{ji} * Incid_i) * H_j + \beta X_j + \sum_{t=2}^{12} \lambda_t Month_t) * Pop_j$$

Here, X_j represents a vector of all the covariates included in the initial equation modeling prevalence. Movement is measured based on the sum of the number of people that enter j from each district i, multiplied by the incidence in district i in 2012, which helps to account for whether people are entering from high or low malaria districts.This variable is also interacted with whether the healthpost catchment area is in a high malaria district in order to test if the effect from movement differs in the two types of districts in the north. In addition, in order to more explicitly model movement from areas with different levels of malaria, a second model is estimated

[3]Extensive tests were done comparing Poisson to Negative Binomial as well as to Zero Inflated and gamma distributions, and the Negative Binomial fit the data best. Results from these tests can be provided upon request.

where the number of people entering are subdivided into five groups based on the malaria prevalence of the district they are coming from, and are included in the equation as five different variables:

$$\text{Cases}_j = exp(\sum_{v=1}^{5} \alpha_v [\sum_{i_v} Enter_{ji}] + \beta X_j + \sum_{t=2}^{12} \lambda_t Month_t) * Pop_j$$

where v is the level of malaria defined as low (1-5 cases per 1000), low-mid (5-15 cases per 1000), mid (15-25 cases per 1000), mid-high (25-50 cases per 1000) and high (50+ cases per 1000), and the number of individuals entering from all districts i that fit category v are summed together. A final model is estimated where these five variables are interacted with the dummy for a high district in order to test for differences in the effect of movement from different areas depending on the existing level of malaria in the district.

5 Results

Results from the simple model of malaria prevalence are presented in table 1 as odds ratios. As would be expected, each additional variable helps to improve the model a little bit, increasing the R^2. The largest jump in goodness of fit is at the addition of the variable for whether the catchment area is in a high malaria district as well as the interaction term. The table shows that for the high malaria areas, rainfall is positively correlated as compared to the low malaria areas, which implies that this environmental factor does not play as large a role in malaria prevalence. The table also shows that the coefficients on the months are very significant and large, implying that the seasonality component is extremely important in explaining the malaria prevalence. Since much of this seasonality variation is due to variation in rainfall, this is explicitly modeled in the appendix using different lags in rainfall.

Focusing now on just 2013, table 2 first shows the same specification in column one as in the last column of table 1, but now it is limited to 2013. Again, we see that rainfall is only significant for the high malaria districts and not for the low malaria ones, where rainfall does not seem to have an effect on malaria prevalence. Including the variable for people entering by incidence rate, the goodness of fit improves for the model. The results in Column 2 also show that the impact of people entering is almost entirely for low malaria districts, while in the high malaria districts the effect of people entering is extremely small. This, combined with the results from rainfall seem to suggest that in areas where malaria is still high, environmental factors play an important role in propagating the disease, while in those areas that have transitioned to low malaria, environmental factors are decreasing in importance and instead the population movement becomes more important for malaria prevalence.

The coefficient on the movement variable is a bit difficult to interpret because it is not only the number of people entering but the combination of number of people and incidence in the place they are coming from, so it is possible to have a place with very few people but high incidence and a place with low incidence but many people, and it is not possible to pull those apart in order to provide a direct policy prescription. Therefore, in the last two columns, instead of using one variable that combines incidence and movement, five variables are used that represent the number of people entering from areas with different levels of malaria incidence, going from low to high as described in the earlier section. The results show that the effect only seems to come from individuals entering a district from the highest malaria areas with over 50 cases of malaria per 1000. This suggests that any policies that aim to target the negative effect of movement on malaria prevalence need to focus on high malaria districts and their relationship to the districts in the north. When the movement variables are interacted with the variable for high prevalence, we see in column four that the effect of people entering from the highest malaria districts impacts both the low and high malaria areas in the North of the country. Interestingly, we also see a negative coefficient associated with people entering the high malaria northern areas from low malaria areas, which could actually indicate that individuals in low malaria areas are avoiding traveling to these high malaria areas when malaria incidence is high.

6 Policy Implications and Discussion

It is possible to use the model that includes people entering scaled by incidence in the area they are coming from, to consider some hypothetical scenarios. The best-case scenario is decreasing malaria prevalence among all travelers. The coefficients estimated from the regression are used and a new estimate is plugged in for number of people entering scaled by incidence, where incidence is decreased down to 5 per 1000 for all health districts where it is above that value.

Figure 7 shows the estimates of total malaria cases predicted based on actual incidence compared to predicted malaria with the lower incidence. The top panel is for the low malaria districts, where in the regressions we saw that movement has a much bigger impact on malaria prevalence, while the bottom panel is for the high malaria districts. As expected, among the higher malaria districts there seems to be very little effect if prevalence among travelers were reduced. Even among the low malaria districts, the figure shows that the effect is very concentrated in a couple of districts where there is much more travel and therefore many new cases are generated by those travelers. This is especially the case in Richard Toll and Saint Louis. This suggests that rather than reducing prevalence among all travelers across the board, which might be very costly, it would be more effective to focus in on the districts that receive many travelers and see where those travelers tend to come from. So focusing on Richard Toll and Saint Louis and where travelers are coming from to those areas could be a very cost effective way for decrease prevalence drastically in those districts.

To narrow down which health districts might be most critical as sources of malaria, I run simulations to calculate the decrease in malaria cases if prevalence among travelers in only one district decreased. This can helps us calculate what would the positive externalities be in the north if resources were only focused on reducing prevalence in one district. The simulation is run for each district separately and the drop in malaria cases in low and in high districts is graphed as a percent of total malaria cases annually in Figure 8. The figure shows that there are a few districts that are key for reducing malaria cases. In particular, the district of Touba, where a big pilgrimage takes place that brings close to a million people from around the country and outside of Senegal, is extremely important. When people from places with different levels of immunity come together for something like a pilgrimage, if the mosquitoes that spread malaria are present (which is the case for Touba), then it can lead to many new infections in individuals coming from low malaria places since they lack any immunity to malaria. When these people travel back to their homes, they can then infect additional people, especially people sleeping in the same household, if mosquitoes are again present. These type of gatherings are therefore important to target with policies that could help prevent these transmissions from happening. Additional important districts are Bakel, Goudry, and Kidira, all in the Southeast of the country where malaria prevalence is very high. Tivaoune is also very important and another location of a big pilgrimage, as is Kaolack, an important commercial center with high malaria prevalence.

This type of analysis can be used by policy makers as they determine how to spend limited resources. It is often not feasible to do certain interventions, such as mass drug administration across a whole country, but without knowing which are the most critical areas to target, it is difficult to know the best policy. This study shows that targeting just a few of the health districts from which the most travelers come could have big impacts in reducing malaria across many districts in the north of the country where Senegal is attempting to eliminate malaria. In addition, the effect from travelers is most important for only a few of the districts, particularly those in the north with very low malaria levels already. Therefore, any policies targeting travel should do so strategically since the impact is only going to be felt in some areas, and only travel impacting those areas should be the focus.

At this point, it is important to mention that cell phone data does have some limitations. First, depending on the percent of the population with access to a cell phone, the data might be capturing movement of only a certain portion of the population, such as those that are high income. In 2013 there were 92.93 mobile phone subscriptions per 100 inhabitants in Senegal, implying that a majority of the population was using cell phones, which would mitigate bias arising from heterogeneous phone ownership (Union 2013). A second concern is that using data from only one mobile provider could lead to a bias if the carrier type is associated with certain characteristics of the user. Out of the three telecom providers in Senegal, Sonatel had between 56 and 62 percent of the cell phone market in 2013 (Régulation des Télécommunications et des Postes 2013). In addition, many individuals have phone cards from more than one provider in order to maximize the promotions of different providers. Based on the "Listening to Senegal Survey" done in 2014, Sonatel is the main provider for 83 % of those surveyed with a cell phone, and 89% of those with a cell phone have a Sonatel SIM card (Agence Nationale de la Statistique et de la Démographie - Ministére de l'Economie 2014). Therefore, while it is important to recognize the limitations of cell phone data, in this study their role is restricted, and in a context with limited data on internal movement, cell phone data provides an opportunity to study short term effects of movement within a country that would not otherwise be possible.

7 Conclusion

The paper set forth to study the direct link between short term population movement and disease prevalence in the context of malaria in Senegal. This type of study has not been possible before due to either a lack of comprehensive data on short term movement at the scale of a country, or else a dearth of detailed data on disease prevalence. New "big data" collected by companies such as cell phone providers is now making the measurement of short term movement possible. While this data still has limitations, in the context of Senegal, the cell phone data is very comprehensive and covers a large majority of the population. In addition, due to the initiative taken by the Ministry of Health to work on eradicating malaria completely from the north of the country, surveillance data has been collected on a monthly level for each health post in the North, which makes it possible to study how short term movement into the areas covered by these health posts can lead to higher prevalence of malaria.

The study finds that there is a significant effect of movement on malaria prevalence, but what is most interesting is the indication that there is a transition that occurs in the mechanisms that propagate the disease. In places with higher malaria, environmental factors such as rainfall play a much bigger role in the prevalence of malaria. As the prevalence is decreased through various policies, the environment of the district is no longer as important in determining malaria prevalence, but the movement of individuals becomes a more prominent mechanism that leads to increased malaria in these very low malaria districts. While this was suspected to be the case, up to now there is no research that has shown the presence of such a transition. With different mechanisms in play in different districts, it becomes necessary for policy makers to consider what types of interventions would be most appropriate for those districts and to target the correct interventions that would be most effective.

Through simulations based on the models predicted, the paper also points to which districts are the most important sources of malaria for the North of the country, as well as which districts in the north are most affected by travel and population movement. Again, this can be used by policy makers to implement more targeted campaigns that would allow the same amount of resources to lead to higher decreases in malaria prevalence.

While the work here focuses on malaria, it is possible to implement these types of models for other diseases as well. As cell phone usage has become extremely prevalent throughout the developing world and cell phone providers are beginning to understand how the data they collect can be used by policy makers to implement better policy, measuring short term movement becomes much easier. What will then become necessary is the collection of high frequency detailed data on all infectious diseases. This data will make it possible to study these types of models that help policy makers better target interventions, and in turn will make it easier to evaluate the interventions if case data is already being collected. This could help countries more effectively fight existing infectious diseases and prevent epidemics of new diseases.

References

Adda, Jérôme (2015). "Economic Activity and the Spread of Viral Diseases: Evidence from High Frequency Data". In:

Agence Nationale de la Statistique et de la Démographie - Ministére de l'Economie, des Finances et du Plan (2014). *Enquete a l'ecoute du Senegal 2014*. Tech. rep.

Balcan, Duygu et al. (2009). "Multiscale mobility networks and the spatial spreading of infectious diseases". In: *Proceedings of the National Academy of Sciences* 106.51, pp. 21484–21489.

Bulletin de Surveillance Sentinelle du Paludisme No 1-46, 2013 (2013). Tech. rep. PNLP. URL: http://www.pnlp.sn/Bulletin-Surveillance-Palu/.

Chabot-Couture, Guillaume, Karima Nigmatulina, and Philip Eckhoff (2014). "An Environmental Data Set for Vector-Borne Disease Modeling and Epidemiology". In: *PloS one* 9.4, e94741.

Chuma, Jane M, Michael Thiede, and Catherine S Molyneux (2006). "Rethinking the economic costs of malaria at the household level: evidence from applying a new analytical framework in rural Kenya". In: *Malaria Journal* 5.1, p. 76.

Cohen, Justin M et al. (2012). "Malaria resurgence: a systematic review and assessment of its causes". In: *Malar J* 11.1, p. 122.

Deshingkar, Priya and Sven Grimm (2004). "Voluntary internal migration: An update". In: *London: Overseas Development Institute* 44.

Doolan, Denise L, Carlota Dobaño, and J Kevin Baird (2009). "Acquired immunity to malaria". In: *Clinical microbiology reviews* 22.1, pp. 13–36.

Enns, E.A. and J.H. Amuasi (2013). *Human mobility and communication patterns in Cote d'Ivoire: A network perspective for malaria control*. Tech. rep. D4D Challenge 1 Book.

Hoshen, Moshe B and Andrew P Morse (2004). "A weather-driven model of malaria transmission". In: *Malaria Journal* 3.1, p. 32.

Huang, Zhuojie, Andrew J Tatem, et al. (2013). "Global malaria connectivity through air travel". In: *Malar J* 12.1, p. 269.

Johnston, Geoffrey L, David L Smith, and David A Fidock (2013). "Malaria's missing number: calculating the human component of R0 by a within-host mechanistic model of Plasmodium falciparum infection and transmission". In: *PLoS computational biology* 9.4, e1003025.

Killeen, Gerry F, Amanda Ross, and Thomas Smith (2006). "Infectiousness of malaria-endemic human populations to vectors". In: *The American journal of tropical medicine and hygiene* 75.2 suppl, pp. 38–45.

Le Menach, Arnaud et al. (2011). "Travel risk, malaria importation and malaria transmission in Zanzibar". In: *Scientific reports* 1.

Littrell, Megan et al. (2013). "Case investigation and reactive case detection for malaria elimination in northern Senegal". In: *Malar J* 12, p. 331.

Lu, Guangyu et al. (2014). "Malaria outbreaks in China (1990–2013): a systematic review". In: *Malar J* 13, p. 269.

Lynch, Caroline and Cally Roper (2011). "The transit phase of migration: circulation of malaria and its multidrug-resistant forms in Africa". In: *PLoS medicine* 8.5, e1001040.

Osorio, Lyda, Jim Todd, and David J Bradley (2004). "Travel histories as risk factors in the analysis of urban malaria in Colombia". In: *The American journal of tropical medicine and hygiene* 71.4, pp. 380–386.

Oster, Emily (2012). "Routes of Infection: Exports and HIV Incidence in Sub-Saharan Africa". In: *Journal of the European Economic Association* 10.5, pp. 1025–1058.

Prothero, R Mansell (1977). "Disease and mobility: a neglected factor in epidemiology". In: *International Journal of Epidemiology* 6.3, pp. 259–267.

Régulation des Télécommunications et des Postes, Autorité de (2013). *Observatoire de la Telephonie Mobile: Tableau de bord au 31 décembre 2013*. Tech. rep.

Shepard, Donald S et al. (1991). "The economic cost of malaria in Africa." In: *Tropical medicine and parasitology: official organ of Deutsche Tropenmedizinische Gesellschaft and of Deutsche Gesellschaft fur Technische Zusammenarbeit (GTZ)* 42.3, pp. 199–203.

Silal, Sheetal P (2012). "Modeling the relationship between precipitation and malaria incidence in Mpumalanga, South Africa". In: *Malaria Journal* 11.Suppl 1, P127.

Siri, Jose G et al. (2010). "Significance of travel to rural areas as a risk factor for malarial anemia in an urban setting". In: *The American journal of tropical medicine and hygiene* 82.3, pp. 391–397.

Smith, David L and F Ellis McKenzie (2004). "Statics and dynamics of malaria infection in Anopheles mosquitoes". In: *Malaria Journal* 3.1, p. 13.

Stoddard, Steven T et al. (2009). "The role of human movement in the transmission of vector-borne pathogens". In: *PLoS neglected tropical diseases* 3.7, e481.

Tatem, Andrew J, Youliang Qiu, et al. (2009). "The use of mobile phone data for the estimation of the travel patterns and imported Plasmodium falciparum rates among Zanzibar residents". In: *Malar J* 8, p. 287.

Tatem, Andrew J and David L Smith (2010). "International population movements and regional Plasmodium falciparum malaria elimination strategies". In: *Proceedings of the National Academy of Sciences* 107.27, pp. 12222–12227.

Union, International Telecommunication (2013). “World Telecommunication/ICT Indicators Database”. URL: http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx.

Wesolowski, Amy et al. (2012). “Quantifying the impact of human mobility on malaria”. In: *Science* 338.6104, pp. 267–270.

Wiser, Mark (2010). *Protozoa and human disease*. Garland Science.

World Malaria Report (2014). Tech. rep. World Health Organization.

Yukich, Joshua O et al. (2013). “Travel history and malaria infection risk in a low-transmission setting in Ethiopia: a case control study”. In: *Malar J* 12, p. 33.

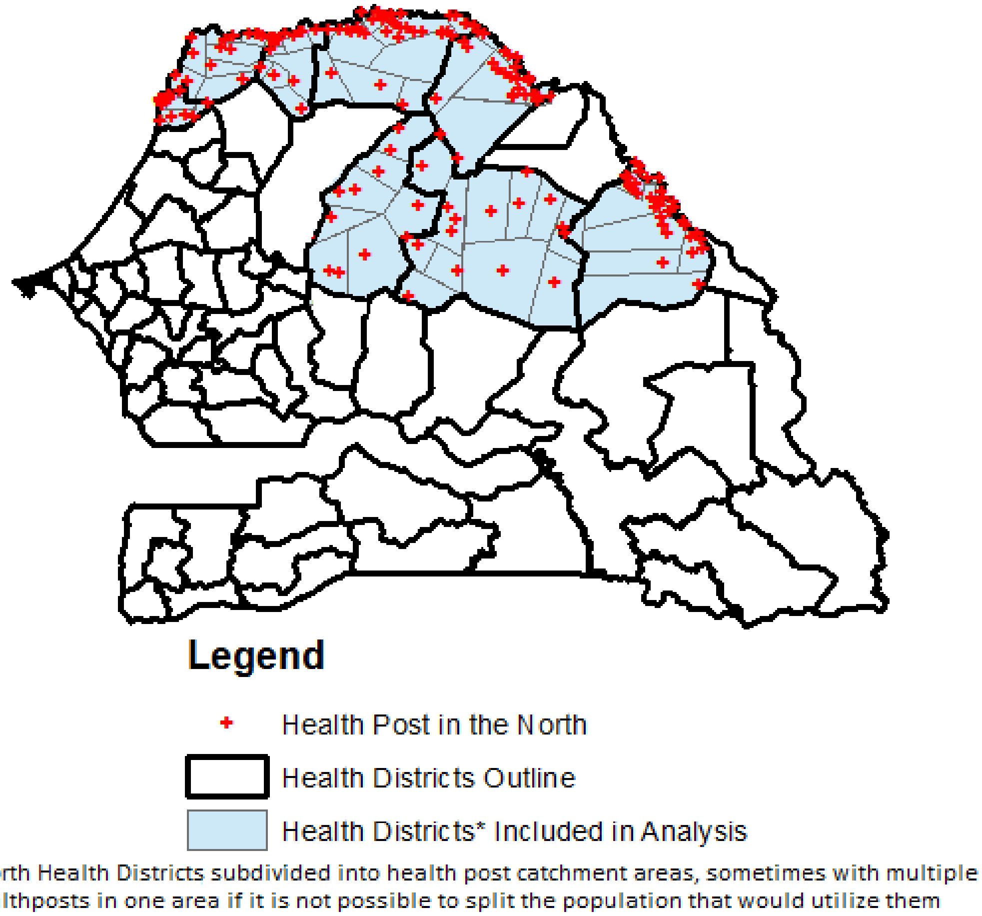


Figure 1: Senegal Health Districts and Location of Health Posts in the North for Which There is Monthly Data

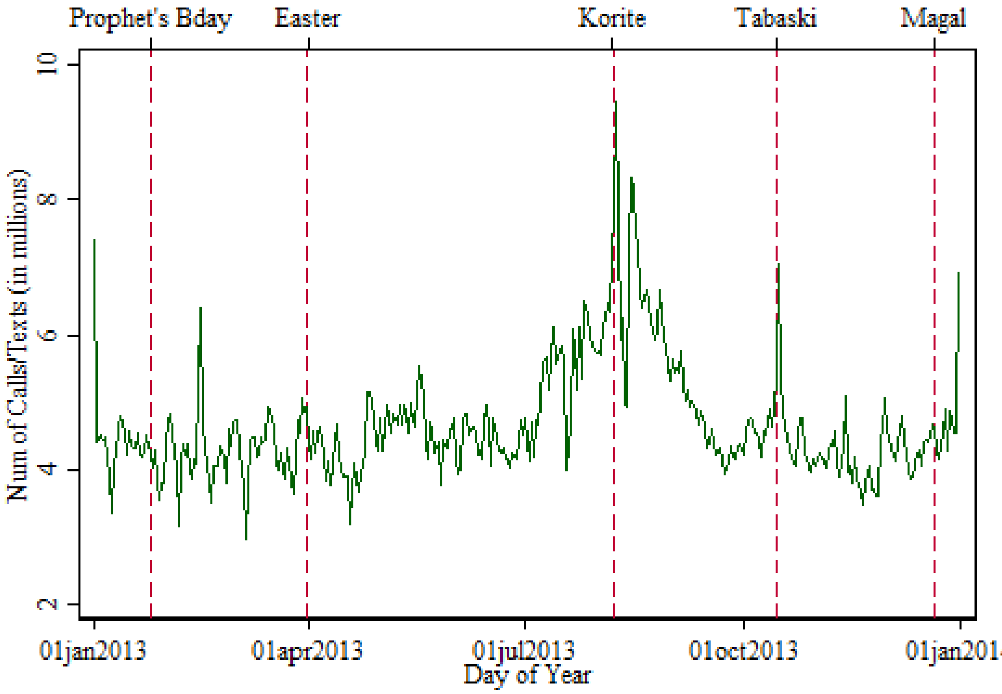


Figure 2: Number of Calls Made and Texts Sent per Day by all Sonatel Users

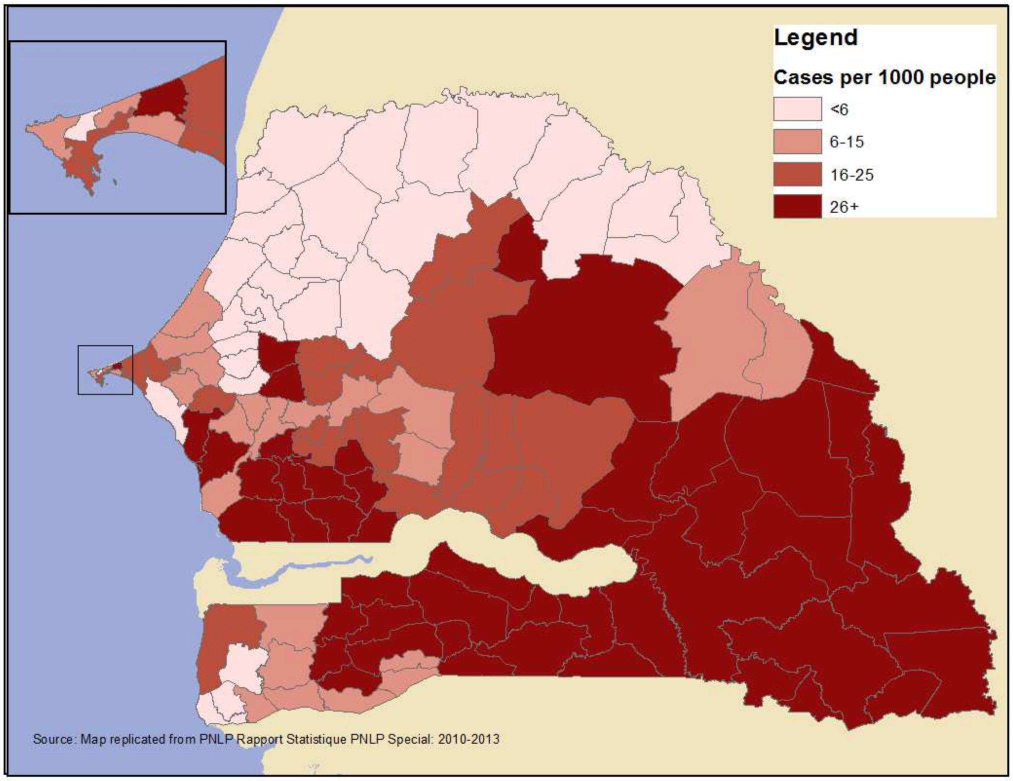
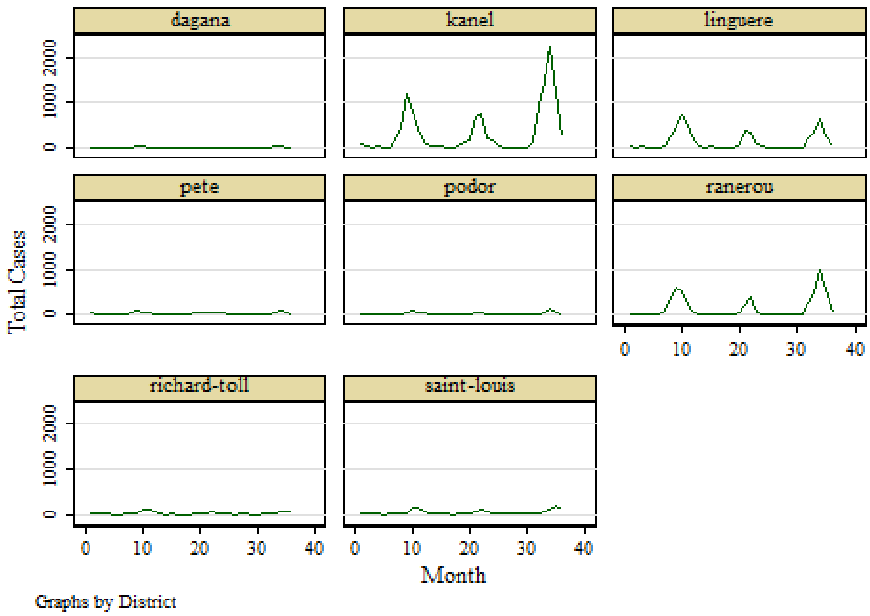
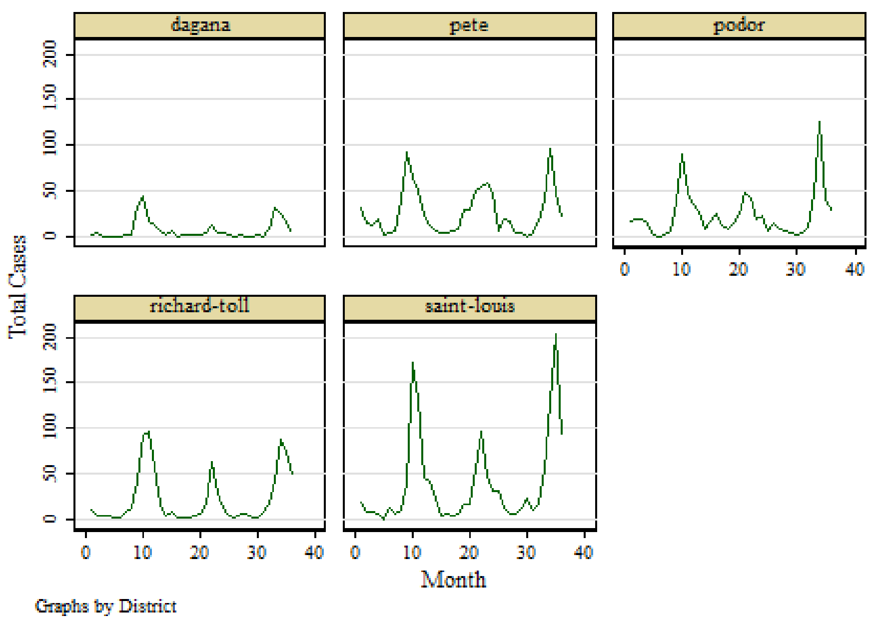


Figure 3: Map of Malaria Prevalence in Senegal in 2013



(a) All North Districts



(b) Low Malaria North Districts

Figure 4: Total Cases per Month

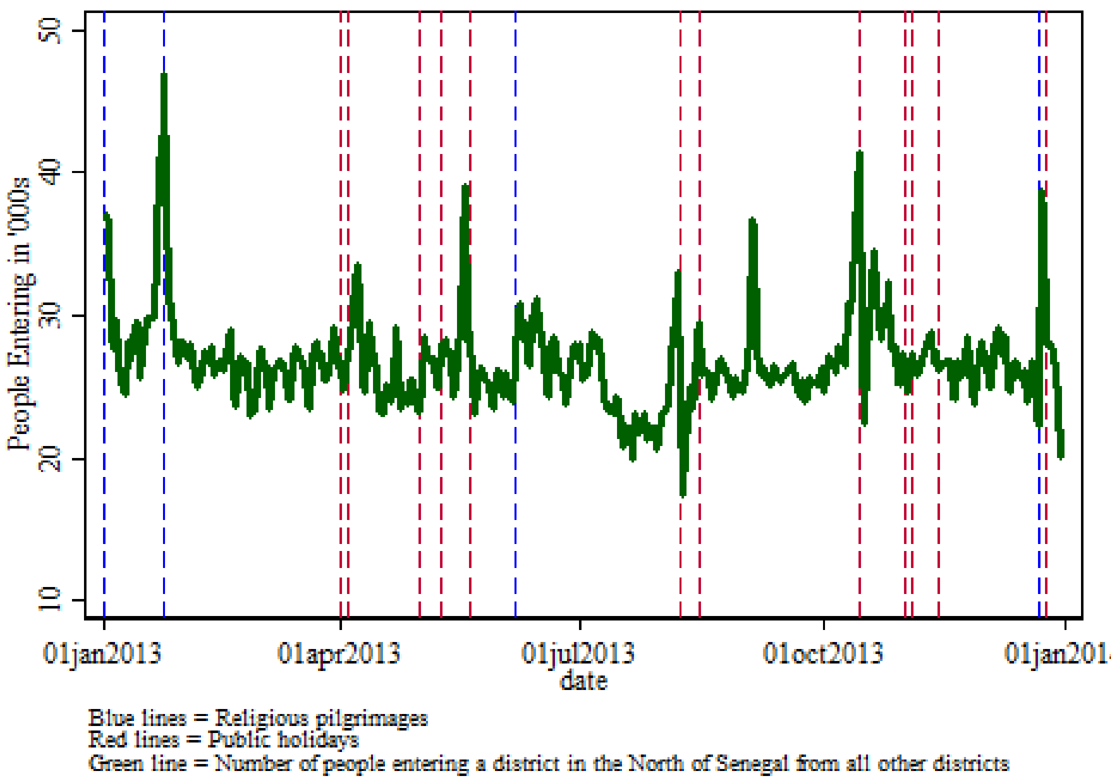


Figure 5: People Entering Northern Senegal vs Holidays and Pilgrimages

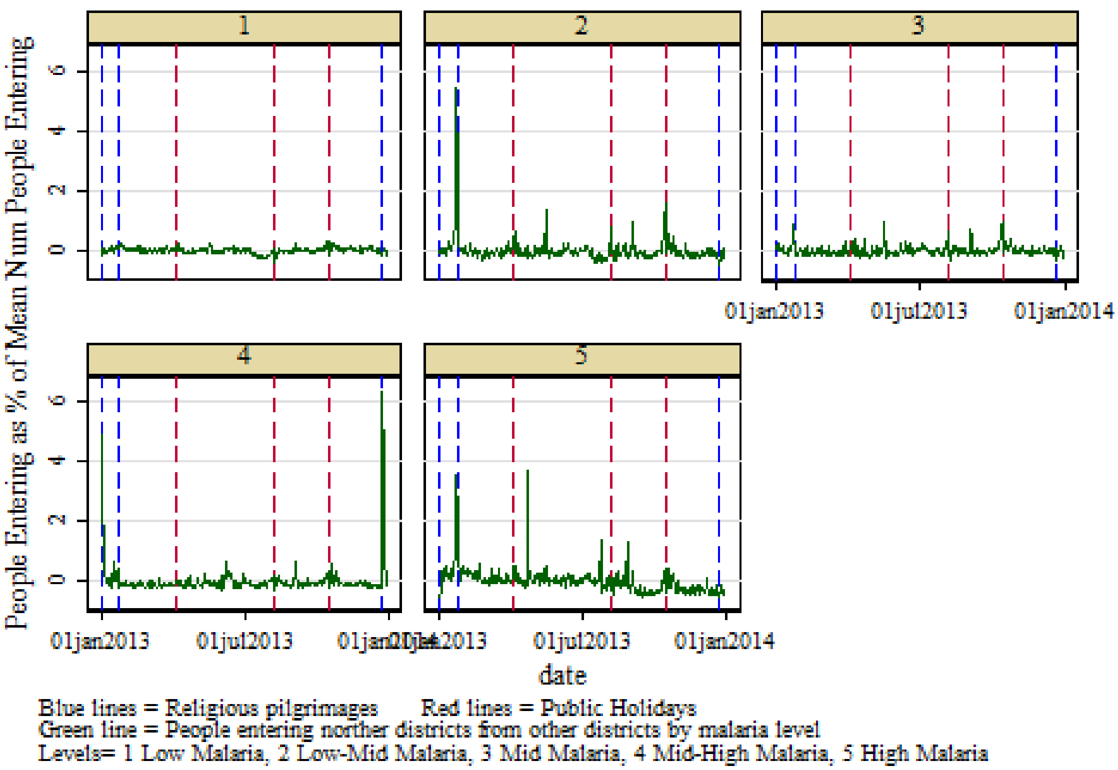


Figure 6: People Entering as % of Mean by Malaria Level

Table 1: Modelling Malaria Prevalence, Excluding Movement

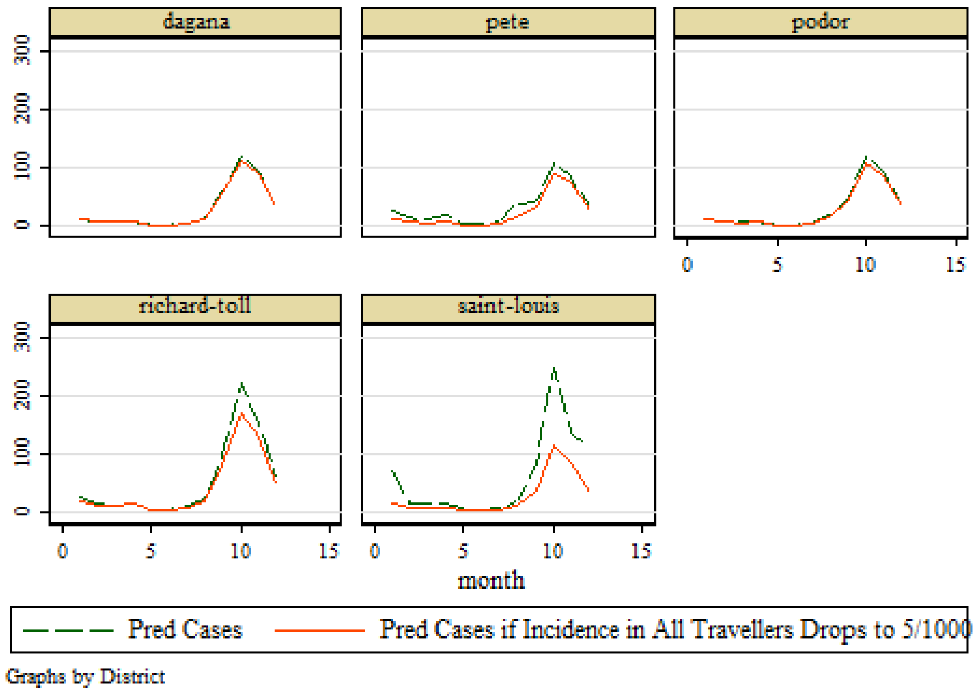
	(1)	(2)	(3)	(4)
Total Rain		1.022***	1.021***	0.999
		(0.00200)	(0.00201)	(0.00279)
Dist to Road			0.928***	0.953***
			(0.0105)	(0.0104)
Rain x High				1.015***
				(0.00194)
High				1.649***
				(0.188)
Feb	0.599**	0.644**	0.629**	0.605**
	(0.124)	(0.137)	(0.140)	(0.138)
March	0.549***	0.568***	0.564***	0.542***
	(0.111)	(0.115)	(0.120)	(0.118)
Aptil	0.442***	0.465***	0.479***	0.458***
	(0.103)	(0.107)	(0.118)	(0.116)
May	0.296***	0.298***	0.293***	0.274***
	(0.0697)	(0.0687)	(0.0694)	(0.0647)
June	0.325***	0.258***	0.254***	0.271***
	(0.0854)	(0.0763)	(0.0744)	(0.0825)
July	1.256	0.267***	0.276***	0.511**
	(0.250)	(0.0695)	(0.0718)	(0.140)
Aug	6.041***	0.345***	0.384***	1.359
	(1.103)	(0.109)	(0.121)	(0.460)
Sept	13.11***	1.097	1.219	3.522***
	(2.339)	(0.315)	(0.372)	(1.004)
Oct	16.77***	5.437***	5.616***	7.855***
	(2.937)	(0.995)	(1.068)	(1.523)
Nov	8.634***	9.168***	8.938***	7.628***
	(1.588)	(1.728)	(1.702)	(1.443)
Dec	2.712***	2.782***	2.758***	2.656***
	(0.476)	(0.500)	(0.498)	(0.500)
2014	0.873	1.024	1.026	0.974
	(0.0887)	(0.102)	(0.107)	(0.105)
2015	0.962	0.803**	0.789**	0.887
	(0.0968)	(0.0862)	(0.0841)	(0.0959)
Constant	1.20e-05***	1.11e-05***	1.33e-05***	1.01e-05***
	(1.67e-06)	(1.55e-06)	(1.93e-06)	(1.69e-06)
McFadden's R^2	0.081	0.093	0.096	0.110
Maximum Likelihood R^2	0.355	0.395	0.407	0.450
BIC	-6060.985	-6199.308	-6236.289	-6390
Observations	2,268	2,268	2,268	2,268

Robust Standard Errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

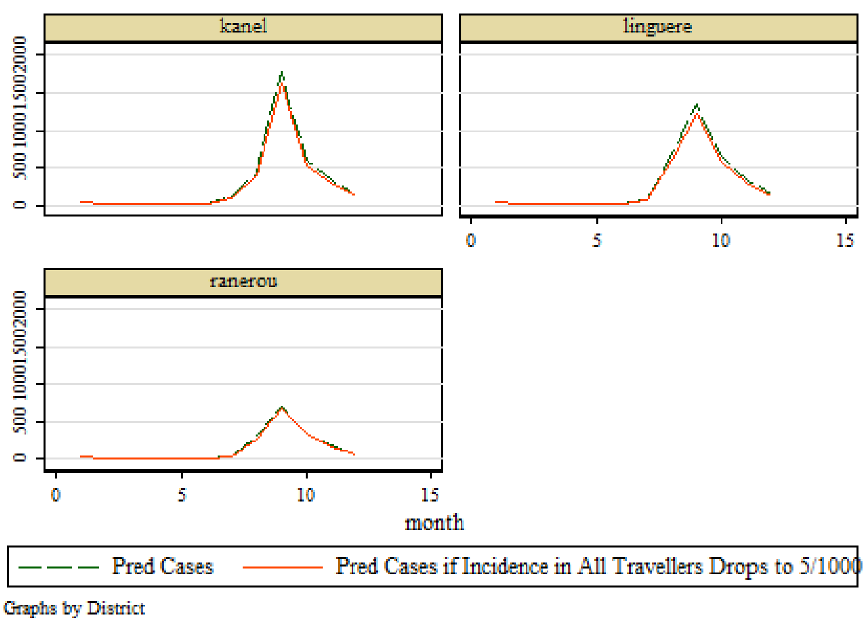
Table 2: Modelling Malaria Prevalence Including Population Movement Scaled by Malaria Incidence of Area Coming From

	(1)	(2)	(3)	(4)
Enter Scaled by Incidence		1.003***		
		(0.000592)		
Enter Scaled by Incidence x High		0.998***		
		(0.000635)		
Enter Low			1.000	1.000
			7.77e-06)	(8.84e-06)
Enter Low-Mid			1.000	1.000
			(1.78e-05)	(2.32e-05)
Enter Mid			1.000	1.000
			(2.04e-05)	(6.10e-05)
Enter Mid-High			1.000	1.000
			(1.12e-05)	(4.26e-05)
Enter High			1.001***	1.001***
			(0.000206)	(0.000306)
Enter Low x High				1.000**
				(1.63e-05)
Enter Low-Mid x High				1.000
				(6.10e-05)
Enter Mid x High				1.000
				(6.54e-05)
Enter Mid-High x High				1.000
				(4.43e-05)
Enter High x High				1.000
				(0.000409)
Total Rain	0.996	0.992	1.026***	0.990*
	(0.00532)	(0.00559)	(0.00451)	(0.00554)
Dist to Road	0.964***	0.991	0.948***	0.987
	(0.0168)	(0.0179)	(0.0177)	(0.0187)
Total Rain x High	1.018***	1.019***		1.018***
	(0.00307)	(0.00325)		(0.00307)
High	1.738***	2.646***		2.474***
	(0.298)	(0.491)		(0.546)
Constant	1.01e-05***	5.49e-06***	1.04e-05***	6.04e-06***
	(2.29e-06)	(1.28e-06)	(2.28e-06)	(1.57e-06)
Month FE	Yes	Yes	Yes	Yes
McFadden's R^2	0.121	0.130	0.115	0.135
Maximum Likelihood R^2	0.493	0.519	0.476	0.531
BIC	-1154	-1182	-1110	-1146
Observations	756	756	756	756

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1



(a) Low Malaria Health Districts in the North



(b) High Malaria Health Districts in the North

Figure 7: Simulation of Malaria Cases if Incidence Drops to 5/1000 Everywhere, Based on Model Including Number of People Entering Scaled by Incidence in Location They Come From

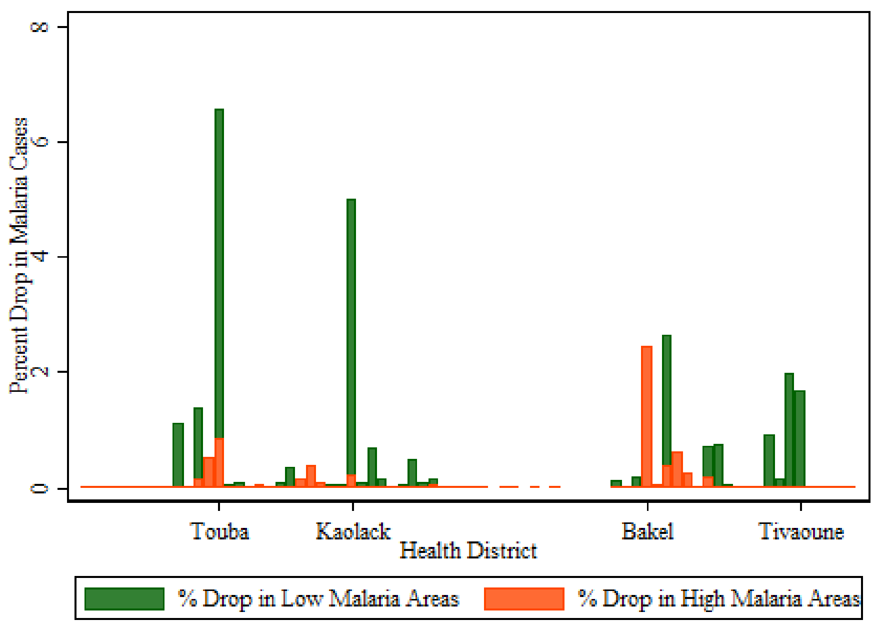
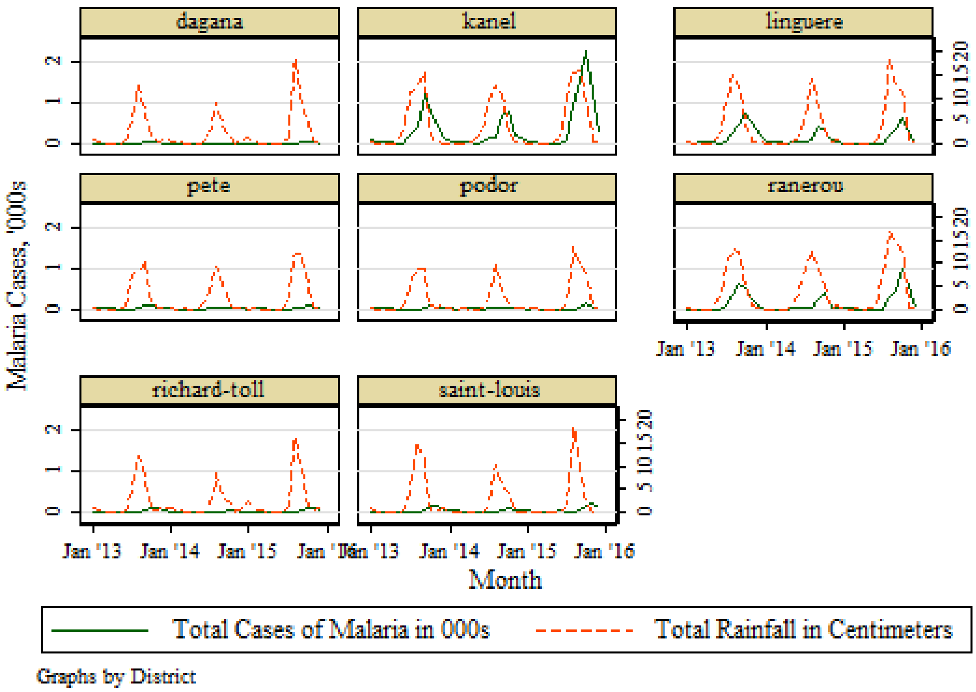
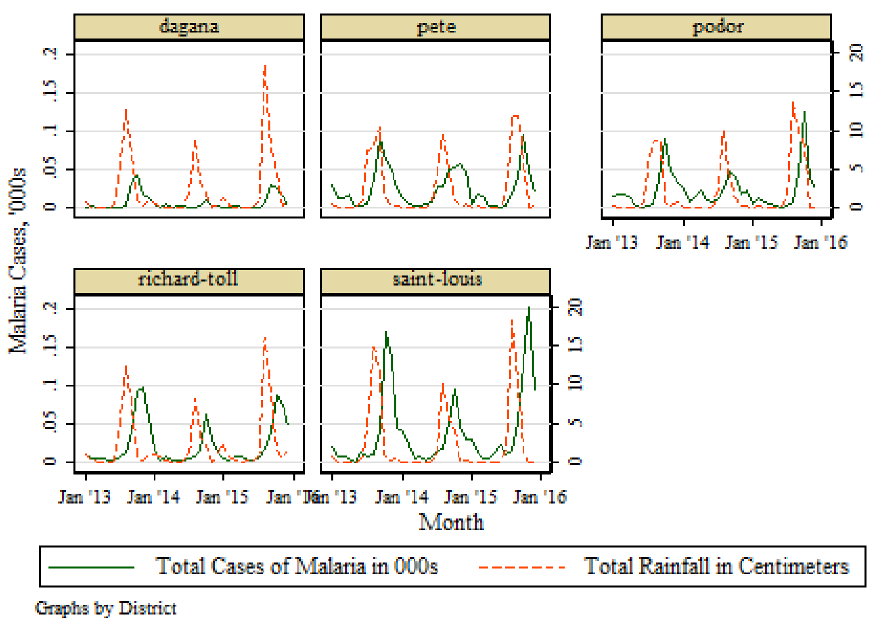


Figure 8: Simulation of Percent Drop in Malaria Cases in Low Malaria and High Malaria Healthpost Catchment Areas if Incidence Drops to 5/1000 in Only One Health District, Result Graphed for Each Health District

Appendices



(a) All Malaria Health Districts in the North



(b) Low Malaria Health Districts in the North

Figure 9: Malaria Cases versus Rainfall, 2013-2015

Table 3: Modelling Malaria Prevalence, Excluding Movement, Seasonality Modeled with Rainfall

	(1)	(2)	(3)
Total Rain	1.009***	1.008***	1.002
	(0.00107)	(0.00108)	(0.00191)
1 Lag Total Rain	1.007***	1.007***	1.005***
	(0.00140)	(0.00143)	(0.00194)
2 Lags Total Rain	1.018***	1.017***	1.016***
	(0.00120)	(0.00130)	(0.00217)
Dist to Main Road		1.004***	0.999
		(0.000935)	(0.00143)
Total Rain x High			1.008***
			(0.00226)
1 Lag Total Rain x High			1.002
			(0.00252)
2 Lags Total Rain x High			1.001
			(0.00250)
High			1.387**
			(0.231)
Constant	6.71e-06***	5.73e-06***	6.51e-06***
	(4.41e-07)	(4.89e-07)	(6.95e-07)
McFadden's R^2	0.090	0.093	0.097
Maximum Likelihood R^2	0.388	0.395	0.411
BIC	-6271	-6292	-6320
Observations	2,268	2,268	2,268

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 4: Malaria Prevalence Including Movement Scaled by Malaria Incidence of Area Coming From, Seasonality Modeled with Rainfall

	(1)	(2)	(3)	(4)
Enter Scaled by Incidence		1.002***		
		(0.000549)		
Enter Scaled by Incidence x High		0.998***		
		(0.000657)		
Enter Low			1.000	1.000
			(8.38e-06)	(9.30e-06)
Enter Low-Mid			1.000	1.000
			(2.19e-05)	(2.27e-05)
Enter Mid			1.000	1.000
			(2.36e-05)	(6.48e-05)
Enter Mid-High			1.000	1.000
			(1.58e-05)	(0.000101)
Enter High			1.001***	1.001**
			(0.000227)	(0.000273)
Enter Low x High				1.000***
				(2.28e-05)
Enter Low-Mid x High				1.000
				(8.40e-05)
Enter Mid x High				1.000
				(7.13e-05)
Enter Mid-High x High				1.000
				(0.000102)
Enter High x High				1.001
				(0.000523)
Total Rain	0.999	0.998	0.998	0.999
	(0.00353)	(0.00377)	(0.00364)	(0.00358)
1 Lag Total Rain	1.005	1.006	1.005	1.006
	(0.00400)	(0.00414)	(0.00400)	(0.00390)
2 Lags Total Rain	1.019***	1.018***	1.020***	1.020***
	(0.00357)	(0.00377)	(0.00363)	(0.00359)
Dist to Road	0.998	0.999	0.996*	0.991***
	(0.00210)	(0.00205)	(0.00248)	(0.00323)
Total Rain x High	1.013***	1.014***	1.014***	1.013***
	(0.00412)	(0.00434)	(0.00427)	(0.00416)
1 Lag Total Rain x High	1.000	0.999	1.000	0.999
	(0.00460)	(0.00475)	(0.00471)	(0.00453)
2 Lags Total Rain x High	1.001	1.001	1.000	1.000
	(0.00399)	(0.00418)	(0.00411)	(0.00400)
High	1.682**	2.183***	2.168***	4.017***
	(0.410)	(0.539)	(0.569)	(1.322)
Constant	6.67e-06***	4.46e-06***	5.10e-06***	5.12e-06***
	(1.05e-06)	(7.51e-07)	(9.40e-07)	(9.86e-07)
McFadden's R^2	0.098	0.106	0.106	0.109
Maximum Likelihood R^2	0.423	0.449	0.450	0.458
BIC	-1110	-1131	-1113	-1091
Observations	756	756	756	756

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$