Less Bite for Your Buck: Using Cell Phone Data to Target Disease Prevention*

Sveta Milusheva

November 18, 2016

Job Market Paper
[Latest Version]

Abstract

Infectious disease has a large economic and social burden that is magnified by infected travelers who spread diseases to the locations they visit. In this paper, I study malaria transmission in Senegal by quantifying the relationship between travel and spread of disease and showing its implication for targeting policies. Using individual mobile phone records for 9 million users in 2013, I estimate daily interregional movement and demonstrate substantial intertemporal and geographic variation in movement. I link this variation to clinic data on the incidence of malaria in order to calculate the probability a traveler is infected and to determine the impact in the area a traveler enters. Estimates indicate that an infected traveler entering a health facility's catchment area causes reporting of 1.6 additional cases. I apply the results to evaluate the potential for policies targeting travelers. At the same cost, strategic targeting of travelers from high-incidence locations would result in up to five times as many cases being averted as compared to current policies of randomly targeting travelers during the malarial season. These findings indicate how novel applications of big data combined with traditional health measures can enable improvements in policy to address negative spillovers from travel and lower the burden from communicable disease.

*I am grateful to Andrew Foster, Jesse Shapiro, and Daniel Bjorkegren for their continuous guidance and support on this project. In addition, I would like to thank Andrew Tatem, Nick Ruktanonchai, Emily Oster, Bryce Steinberg and Victor Alegana for valuable feedback. I also appreciate the helpful comments from participants at the World Bank ABCDE Conference, the 2016 NEUDC Conference, the 2016 PopPov Conference, the 2016 Population Health Sciences Workshop, the PODER Summer School, the 2016 PAA Conference and the Brown microeconomics lunch seminar. Thank you to the Bill & Melinda Gates Foundation and the National Institute for Child Health and Human Development (Grant # T32 HD 7338-29) for financial support. Finally, this research would not have been possible without the data provided by Orange and Sonatel's Data for Development Challenge, the National Malaria Control Program and MACEPA.

1 Introduction

The Global Fund has disbursed nearly $28.4 billion in the last decade to reduce the disease burden from malaria, TB and HIV since malaria elimination and elimination of other infectious diseases have been shown to have important positive economic consequences (The Global Fund, 2016,Bleakley, 2007, Bleakley, 2010). However, travelers can reverse the progress from campaigns that have decreased infectious disease prevalence (Cohen, D. L. Smith, et al., 2012, Lu et al., 2014), or can rapidly spread emerging diseases such as Ebola and Zika (Tam, Khan, and Legido-Quigley, 2016, Bogoch et al., 2016). Spillovers can arise when infected individuals from regions with a higher disease burden enter areas where a disease has been eliminated or reduced. For example, Venezuela, the first country certified by the WHO for eliminating malaria, has seen a large resurgence of the disease in the last year due to migrant workers in the mining region who have become sick, traveled home, and spread the disease to their home villages and cities (Casey, 2016). Movement and migration also have many benefits. How can societies capture these benefits but limit the risks from the associated spread of disease?

The main challenge in studying the impact of travel on disease and how to mitigate it is the difficulty of tracking population movement and infections. Both disease transmission and movement are highly variable over space and time; therefore, detailed data are necessary to account for their seasonality and spatial distribution. For example, if a disease is highly dependent on environmental factors like rainfall and is only present during the rainy season, annual data on population movement and disease could provide a distorted estimate of the relationship between the two. Additionally, even if the measurement challenge is overcome and the relationship between travel and disease is quantified, it is unclear how a government should determine who are the most cost effective travelers to target in order to prevent the spread of diseases.

This paper overcomes this challenge by utilizing a new source of data—mobile phone

records—to track population movement for a large number of people at a very high spatial and temporal resolution. I combine the mobility data with comprehensive disease incidence data. In this paper, I employ an empirical model of the relationship between population movement and disease to measure the effect of a single infected traveler to an area on the area's disease burden. The model implies that the risk posed by a traveler is a function of the traveler's origin, destination and time of trip. The empirical model suggests a cost-effective strategy to target travelers with preventative interventions. I find that for a given government budget, switching to the cost-effective strategy can avert up to 60 percent more cases as compared to the next best strategy.

To quantify the impact of travel on malaria cases, the paper builds on an existing epidemiological model of malaria propagation to incorporate population movement and separate locally generated and imported cases. I use mobile phone data for 9 million SIM cards in Senegal in 2013 to extract patterns of movement between different areas from the approximate location of 15 billion calls and texts. From the origins, time and length of travel of incoming travelers to an area, I compute the expected number of imported malaria cases. I use a panel data strategy in which I estimate the impact of imported incidence on malaria incidence controlling for location fixed effects and time dummies. If infected travelers only represented a change in location of a case, but did not generate any externality in the form of additional infections, then for each expected imported case there should be just one more additional case reported in the area. Instead, I find that one additional expected imported case of malaria entering a low malaria area leads to 1.6 cases of malaria reported in that area; therefore, contributing to the propagation of the disease. I examine the impact of expected imported cases both in areas that are close to eliminating malaria and ones with a high case load, and I find that the impact of imported cases is significant in the areas close to elimination. The results for the areas with high malaria incidence are inconclusive.

There are several potential threats to identification that I address in the paper. I find

that population movement over time within an area is largely driven by agricultural seasons and holidays, so I consider confounders that might be correlated with these factors and with malaria. I control for the agricultural seasons in my analysis using rainfall. I use a placebo test to ensure the results are not being driven by holidays increasing movement and also causing malaria to increase through another channel such as individuals spending more time outside during holidays. Additional placebo tests are used to check that the relationship between imported cases and malaria incidence is not being driven by some other relationship between the location where imported cases are coming from and the location they enter as well as to ensure that the relationship only holds for expected imported cases of malaria and not other health conditions. I also test whether future imported cases impact malaria incidence in the current month and find that they do not. I argue that this evidence is suggestive of a causal relationship.

This paper demonstrates how detailed movement data can be used to craft a targeted intervention that addresses negative externalities associates with travel. I study four strategies for addressing the impact of travelers. The simplest is a random untargeted strategy throughout the year, the second is a random strategy during the malaria season, the third targets districts based on malaria incidence in the prior year, and the final targets travelers in certain months using my estimated mobility and incidence model. I run simulations based on the model estimated and construct four cost-curves for the potential targeting strategies. By construction, the cost-effective strategy based on my model leads to the largest decrease in cases deriving from travelers for the same number of travelers targeted. Yet, the size of the gain ranges from being 30 percent more effective to being over five times as effective depending on the alternative strategy used as a comparison.

This paper relates closely to work by Tatem, Qiu, et al., (2009) and Le Menach et al., (2011). They use movement patterns from three months of cell phone data and endemicity data to estimate the malaria importation rate to Zanzibar as 1.6 cases per 1000 residents per year. I build on this work by developing an epidemio-

logical model that I can empirically estimate using incidence data[1] along with a full year of mobility data that allows me to capture the seasonality and spatial variation in movement and malaria prevalence. Annually, I find an importation rate of only 0.58 cases per 1000 residents. When this is disaggregated monthly, the rate during high malaria months is 38 times the rate in low malaria months.

More generally, the paper relates to the literature that establishes the relationship between population movement and malaria and finds that travel is an important risk factor for contracting malaria (Lynch et al., 2015, Osorio, Todd, and Bradley, 2004, Siri et al., 2010, Yukich et al., 2013, Alkhalife, 2003, Littrell et al., 2013). This paper shows also the negative externality of travel in the form of secondary cases. Thus far, prior work using mobile phone data to study population movement and malaria has categorize the transmission risk for a set of origins and destinations, but has stopped short of calculating the number of new cases due to travel patterns (Wesolowski, Eagle, et al., (2012) and Enns and Amuasi, (2013)).

The policy simulations of the paper relate to the literature on cost-effectiveness of health interventions (Canning, 2006) and specifically, reducing the negative health spillovers of travel. Previous work in this area has studied nationwide policies (Adda, 2016, Lima et al., 2015), focused on a particular mode of transit (Epstein et al., 2007, Adda, 2016), or else has considered hotspots of disease without incorporating the cost-effectiveness of targeting particular hotspots (Bousema et al., 2012, Cohen, Dlamini, et al., 2013, Wesolowski, Eagle, et al., 2012). This paper uses the estimated model of the relationship between population movement and malaria for a quantitative policy design exercise that evaluates the cost-effectiveness of targeting specific groups of travelers with policies.

[1]The endemicity data are gathered from parasite rate surveys in which a random subsample of the population is tested for malaria parasites. When the malaria prevalence is very low, though, the likelihood of having a positive case becomes very small. Therefore, incidence (new cases tested at a health facility) can be a more reliable measure in low malaria settings (Alegana et al., 2013, Cohen, Dlamini, et al., 2013)

A plethora of research has shown that as infected individuals travel, they can carry disease with them and transmit it in new areas, leading to negative spillovers from areas where the disease is prevalent (Oster, 2012,Prothero, 1977, Balcan et al., 2009, Huang, Tatem, et al., 2013, Tatem and D. L. Smith, 2010, Wesolowski, Metcalf, et al., 2015, Wesolowski, Qureshi, et al., 2015, Stuckler et al., 2011, Pison et al., 1993). Yet, more work is necessary to understand how to manage these spillovers. Given the billions of dollars spent on reducing infectious diseases, it is critical that the money is spent effectively in order to prevent reversing the progress made. While this paper shows how a policy targeted at the most cost-effective travelers can lead to a large reduction in imported malaria cases, it also has implications for other diseases, since the travel patterns studied using the cell phone data could lead to the transmission of any communicable disease. If these data are obtained for other countries or for different diseases, it is possible to replicate the analysis using the methodology developed in this paper. I demonstrate how new sources of big data can be used to study externalities associated with population movement and can help to inform policies to fight infectious disease.

The paper begins by providing some background and describing the data. It then goes on to model the link between malaria and population movement in section 3. Section 4 outlines the empirical results linking travel to malaria and section 5 examines the cost effectiveness of four different policies. Some robustness checks are provided in section 6, and the paper concludes with section 7.

2 Background and Data

2.1 Malaria Characteristics

Malaria is an infectious disease that requires two hosts–humans and mosquitoes–in order to spread. Half of the malaria parasite cycle occurs in the mosquito and half of the cycle occurs in the human, so both are necessary for the spread of the disease. Several malaria parasites exist worldwide, but in Senegal *P. falciparum* is considered

to cause 100 percent of cases (*World Malaria Report* 2014). The malarial cycle for *P. falciparum* can take several weeks. After an infected individual is bit by a mosquito, there is an incubation period lasting around 9 days within the mosquito (Killeen, A. Ross, and T. Smith, 2006).[2] If the mosquito survives the incubation period, it can bite and infect a healthy individual, after which there is a second incubation period within the human of around 15 days (D. L. Smith and McKenzie, 2004, Hoshen and Morse, 2004). Symptoms will appear at the end of this period and the individual will become infectious.[3] This second person is known as a secondary case arising from the first (primary) case. Combining the two incubation periods, a secondary case will take around one month to appear after a primary case. In areas where the disease is highly prevalent, acquired immunity can develop. The person with such immunity still has a small amount of the parasite in them, but does not exhibit any symptoms.[4]

There are two channels through which population movement can spread malaria from high malaria to low malaria areas. The first is residents of low malaria areas who travel to high malaria areas and become infected when bit by infected mosquitoes. Since malaria symptoms do not appear for around two weeks, the resident can travel home feeling healthy. Once at the home location, the person can become symptomatic, as well as infect mosquitoes. These infected mosquitoes can infect other individuals and pass on the disease. The second channel is visitors or migrants that live in a high malaria area and travel to a low malaria area. Again, at the beginning of their travel, these individuals might not exhibit symptoms, but can still be carriers of the disease.[5] Therefore if they are bit by a mosquito in the low malaria area,

[2]The incubation period can vary, but the value of 9 days was found to be the average for two different sites in Senegal.

[3]Unlike some of the other malarial parasites, *P. falciparum* does not have the potential to lie dormant for months or years.

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

[5]There are three possibilities in the case of visitors. 1. The person was recently infected and the incubation period finishes after arrival at a new location, at which point the person becomes symptomatic and can infect mosquitoes. 2. The person has acquired immunity but develops symptoms after arriving to the low malaria area and can then infect mosquitoes. Anecdotally,

they could infect that mosquito and it could in turn infect other individuals in the household or nearby areas.[6]

2.2 Health System and Malaria in Senegal

Senegal is geographically divided into 14 health regions, under which there are 76 health districts. Senegal's health system consists of three tiers: hospital level, health center level, and health post level. A health post usually has four or five health workers that provide preventative and primary curative services and are the first point of service for malaria cases. There are a total of 1,247 health posts in the country (PNLP, INFORM and LSHTM, 2015). In addition, there are rural health points and huts as well as community health workers that provide care for those living far from a health post, and there are mechanisms to facilitate transfer of information between these entities and the health posts.

2.2.1 Malaria Incidence and General Malaria Policies

Since the establishment of the National Malaria Control Program (PNLP) in 1995, the program has coordinated a variety of measures and policies that have led to a reduction in deaths attributed to malaria from 12.93 per 100,000 people in 2000 to 8.26 in 2013 and 4.0 in 2014 (PNLP, INFORM and LSHTM, 2015). Figure 1 shows malaria incidence throughout Senegal in 2013. The north of the country has very low incidence and is at the level considered ready for elimination by the World

people often travel to a new location for work, and the physical stress of the work lowers the immune system causing the parasite that was within the immune person to develop faster and leading to the development of symptoms and a higher likelihood of infecting mosquitoes. 3. The person has acquired immunity and never becomes symptomatic. These people can still infect mosquitoes and lead to new cases. Since the data used in the paper measures only symptomatic individuals that are tested by a health worker, I do not capture these asymptomatic individuals. Nevertheless, research points to the fact that while these individuals can spread malaria, the rate is two to sixteen times lower compared to symptomatic individuals (Okell et al., 2012)

[6]Since the average radius of travel for the mosquitoes that carry the malaria parasite in Senegal is around 1-2 kilometers, the study does not consider the movement of mosquitoes because the analysis is carried out on a larger geographic scale(Russell and Santiago, 1934, Thomas, Cross, and Bøgh, 2013).

Health Organization (1 case per 1000, also known as the pre-elimination phase). In contrast, the south still has a high case load, with some districts as high as 270 cases per 1000. It is considered to be in the control phase, when it is necessary to first lower the disease burden before the area is ready to attempt elimination. The heterogeneity can be partly attributed to environmental factors because the rainy season is twice as long in the south as in the north, which allows for mosquitoes to breed and spread the disease for longer. This difference has made it easier to control the disease in the north, yet, the mosquitoes are still present there and able to spread malaria (Ndiath et al., 2012).[7]

Given the two distinct zones in the country, the government of Senegal follows WHO guidelines for the types of policies that should be implemented in each. In the control area these include vector control, intermittent administration of full treatment courses of an antimalarial medicine to children, use of Rapid Diagnostic Tests (RDTs) and treatment. In the pre-elimination area these are an enhanced surveillance system, detection, treatment and reporting.

2.2.2 Malaria Policies Towards Travelers

MACEPA, a non-profit organization working with PNLP to fight malaria in Senegal, has specifically focused on reducing imported cases in the district Richard Toll. They have used volunteers in the community to alert health workers to the arrival of new travelers. Health workers track down these travelers and ask to test them using an RDT, and treat those that test positive. Based on 2015 data provided by the Richard Toll Health District Director, 3,609 people were identified as travelers, of these 3,386 were tested and 10 tested positive for malaria. In 2015, there were a total of 186 imported cases;[8] therefore, the strategy of testing anyone determined to be a traveler by a volunteer was only able to detect 5% of imported cases.

[7]Construction of dams along rivers in the north in recent years has especially provided breeding sites for the mosquitoes.

[8]Data on imported cases is based on surveys conducted with each individual testing positive for malaria in Richard Toll. The survey asked about travel history to determine if the case was generated locally or imported.

A more systematic policy to target travelers was implemented in one health post in Richard Toll, which is privately run by the Senegalese Sugar Company (CSS) for its workers and their families. CSS hires over 3000 migrant workers every year to help with the sugar harvest. Malaria was a large burden for the company, causing lower productivity, high absenteeism, and high spending on pharmaceuticals to treat malaria ($23,000 spent on malaria pharmaceuticals over 6 months in 2011) (Djibo and Ndiaye, 2013). In late 2011, the CSS implemented a new mandatory policy for all seasonal workers: testing every worker at the beginning of the season using an RDT, treating anyone testing positive, and providing workers and their families with bednets and information.

Figure 2 shows the trend in malaria cases at the CSS health facility from 2008 through the end of 2015 based on data provided by the CSS. The red vertical line marks the implementation of the policy in late 2011. There was a drastic decrease in cases after the implementation of this policy, with case numbers at zero or close to zero in most months after the policy. The figure also compares the malaria data with data on two types of schistosomiasis among workers at the CSS and shows no drop in those diseases in late 2011, demonstrating that the drop in malaria cannot be attributed to an overall improvement in the healthcare facility. The targeted policy aimed at travelers helped to decrease malaria significantly down to where it can be completely eliminated.

A strategy to decrease malaria cases that has proved effective and could be harnessed for targeting travelers is proactive community treatment (ProACT) implemented by trained home care providers (HCPs). The pilot of this intervention consisted of HCPs going door-to-door weekly to every household in a village, checking for individuals with symptoms. Those exhibiting symptoms were tested with an RDT, and individuals with positive tests were either treated or referred to the closest health facility. Compared to villages that did not receive ProACT, the odds of symptomatic malaria were 30 times lower in the intervention villages (Linn et al., 2015). This type of pol-

icy could be applied to areas that receive travelers during certain times of the year. For example, HCPs could be sent out during the weeks when the most potentially infected travelers are coming in, in order to go door-to-door to test travelers. For this type of policy, it is necessary to know when and where HCPs should go, which is difficult to know but is made possible by the analysis that I conduct.

A final policy, which is currently being discussed but has not been implemented, is to utilize mobile phones to target travelers with information.[9] Mobile phones are always linked to a tower if they are turned on; therefore, the provider knows as soon as an individual changes towers. A targeted text message can then be sent as soon as someone changes location from a high malaria tower to a low malaria tower, recommending and providing an incentive to get tested at the closest clinic for free. This type of targeted messaging is already being used for marketing, while text messaging more generally has been used by the Ministry of Health in Senegal for providing information on diabetes.

2.3 Population Movement in Senegal

Senegal has large flows of long term and permanent migration, with 27% of the population recorded as an internal migrant in 2004 (P. D. Fall, Carretero, and M. Y. Sarr, 2010).[10] A large part of this migration is rural to urban due to irregularity of rainfall and degradation of the ecosystems that have impacted agricultural activity and significantly reduced the agricultural production index (P. D. Fall, Carretero, and M. Y. Sarr, 2010, Goldsmith, Gunjal, and Ndarishikanye, 2004). In turn, this longer term migration can lead to commuting patterns and visits as people either return home to visit family and friends or those family and friends come to visit them since long distance travel is in large part influenced by social networks (Cho, Myers, and Leskovec, 2011). Specifically looking at migrants in Dakar, A. S. Fall, (1998)

[9] I am currently in discussions with MACEPA and the mobile phone provider Sonatel concerning the possibility of implementing this policy.

[10] This is comparable to the rest of sub-Saharan Africa, where an estimated 50-80 percent of rural households were cited to have at least one migrant member (Deshingkar and Grimm, 2005).

finds that 87% of male migrants and 81% of female migrants visited their home areas. Fall (1998) finds that the large majority of these visits occur for holidays, family ceremonies and religious festivals.

In addition to long term migration and short term visitation patterns triggered by these migrations, detailed studies of the Jola ethnic group in several villages finds that circular migration plays an important role, with over 80% of unmarried Jola youth traveling to the cities in October and then coming back before the rice harvest in June-July (Linares, 2003). Additionally there are still important pastoral groups such as the Fulani that might travel within a set territory, but may also send paid herders with their livestock over longer distances in order to maintain their herds depending on availability of pastures (Adriansen, 2008). Finally, research on youth in Senegal has shown that more than half of the internal migration they engage in is temporary and rural to rural or urban to urban, not rural-to-urban permanent migration (Herrera and Sahn, 2013).

Understanding the movement patterns within Senegal is important for thinking through potential confounding factors between movement and malaria. The majority of the literature points to movement triggered by agricultural seasons as well as holidays. Both of these factors and their relationship to malaria prevalence will be discussed in the model section.

2.4 Malaria Data

The data used to measure malaria incidence comes from the PNLP and MACEPA, which have collected case data at different spatial and temporal levels. The cases are all new cases in the reporting month. If an individual feels sick, usually experiencing a fever, chills and fatigue in the case of uncomplicated malaria,[11] she will go to the closest health post where she will be tested using a rapid diagnostic test

[11] Almost all cases are uncomplicated, with a few severe malaria cases when infections are complicated by serious organ failures and symptoms include coma, difficulty breathing, and acute kidney failure. These cases need to be taken to a hospital immediately.

because she exhibits symptoms of malaria. If she tests positive, she will be provided with medication for free to treat the disease. The process of testing all potential malaria cases ensures that the data consists of confirmed malaria cases only. Case data are available monthly by health district for the whole country and monthly at the health post level for eight health districts starting in 2012.

The analysis focuses on the eight health districts for which data are available at the health post level, covering 184 health posts. In these districts the government has implemented case investigation and surveillance in 2015, whereby for each confirmed positive malaria case, information is sent to the district health supervisor and within three days of notification, a team is deployed to the community to investigate the household of the confirmed case and the five neighboring households. The next step in eliminating malaria in these districts is to pro-actively target imported malaria cases instead of waiting for them to visit a health post and be tested, at which point they might have already infected other individuals.

There are three main challenges that arise with using clinical data: incomplete data reporting, presumptive diagnosis based on symptoms rather than being diagnosed parasitologically and non-utilization of the public health system (Alegana et al., 2013). In the data only 1.7% of health post-month observations are missing.[12] In addition, 97% of suspected cases were tested parasitologically in the eight health districts with health post level data. Utilization of the public health system is more difficult to discern. Since both malaria cases and imported cases are calculated based on case data, as long as utilization is relatively uniform across the country, the effects measured should still be accurate. Based on the DHS data for all the regions, a health facility was visited for fever in children under age 5 in 46% of cases (ANSD and ICF International, 2015). The standard deviation of this utilization across regions is 6.5 percentage points. Health post utilization throughout the country might actually be higher than what is recorded in the DHS data, though. In a smaller survey at a Demographic Surveillance Site in Fatick, Senegal, Franckel and Lalou,

[12]There are a total of 2,208 health post-month observations.

(2009) find that nearly 75% of respondents consider the most important step to treating uncomplicated malaria to be visiting a health facility. In a different survey of farming households located in three communities within the eight health districts analyzed here, Lépine and Le Nestour, (2012) find that 84% of individuals sought treatment from a health care provider during their last illness. In my analysis, I will show results scaled by regional utilization from the DHS.

Figure 3 shows the health districts for which there is monthly data at the health post level. I subdivide these health districts into areas based on the location of the health posts and cell phone towers. Health posts in close proximity (in the same village or neighboring villages) were grouped together into one catchment area. There are 63 health post catchment areas in total, and the subdivisions are shown on the map in light gray lines. I also categorize the eight health districts into five low malaria districts and three high malaria districts. Low and high malaria districts are different in terms of important environmental factors such as rainfall as well as the level of immunity present in the population;therefore, it might be expected that the impact of travelers could be different between the two. I therefore study the impact of travelers separately for these low and high districts. Low and high groupings are marked in light and dark blue respectively in Figure 3. The average size of the healthpost catchment areas is $309mi^2$.

Monthly malaria incidence per 1000 people in each health post catchment area is averaged for each of the eight health districts for three years in Figure 4. The figure highlights the stark contrast between the low and high malaria districts, shown in the top and bottom panels respectively. While the five low malaria districts on average have around 0.1 cases per 1000 people per month, the high malaria districts average 1.9 cases per 1000 people per month.

In addition, the figure also overlays the monthly cumulative rainfall in centimeters in the health post catchment areas averaged by district. The comparison of cases and rainfall demonstrates strong seasonality of malaria in Senegal and the close rela-

tionship between rainfall and malaria, with the peak of cases annually occuring one to two months after the peak in rainfall. While both low and high malaria districts experience peaks of malaria two months after the peak in rainfall, the relationship between rainfall and malaria cases is stronger for the high malaria districts as compared to the low malaria districts.[13] The relationship between malaria and rainfall is explicitly modeled in the analysis since rainfall is an important explanatory variable of malaria that could also affect population movement and importation of cases from other districts.

2.5 Population Movement Data

The data used to measure short term movement come from phone records made available by Sonatel and Orange in the context of the Data for Development Challenge, a call for projects with the objective to explore the potential of mobile call data to facilitate socio-economic development. The data, which were provided to six teams, come from the second phase of the project and consist of call and text data for Senegal between January 1, 2013 and December 31, 2013 for all of Sonatel's user base.[14] I want to use the data in order to capture all population movement within Senegal, and specifically movement to the eight health districts that the analysis is focused on.

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 14.2 million. While coverage is very high, not every individual is included directly with a SIM card. There are two sets of people that are potentially excluded from the data and need to be accounted for–those without a phone and those with a phone but using a different mobile provider. 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

[13]Partial R^2 of 0.37 for rainfall in the low districts as compared to 0.44 for the high districts.

[14]The first stage of the challenge provided a small sample of the data to researchers, and from the projects submitted, 6 were selected to be implemented and were provided with the full set of 2013 data.

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, I am missing only around 11% of individuals with a cell phone but with a different provider. Individuals without a phone can be split into two groups–adults and children. There were 92.93 mobile phone subscriptions per 100 adults in Senegal (Union, 2013), and based on the 2013 Demographic Health Survey (DHS), 93% of households owned at least one cell phone (ANSD and ICF International, 2015). In addition, based on the Listening to Senegal survey, of those people that do not own a phone, 79% state that they own a SIM card which they use in someone else's phone. Therefore, of the small percent of adults without a phone, the majority are already included in the data because they have a SIM card. This implies that most of those without a phone are children under the age of 15, who make up around 40% of the Senegalese population.

I can weight the mobile phone data in order to account for the individuals not included. Using the DHS survey, it is possible to look at the difference in mobility patterns between women with and without a cell phone in the household. There is no statistical difference in whether a trip longer than a month was taken between those with and without a cell phone. The women with a cell phone in the household have only a slightly higher average number of trips taken in the last year. This suggests that I can weight making the assumption that the movement patterns for those not in my data are similar to those in my data. A big part of the missing group is children, though, who are likely traveling with adults that are represented in the data, but might be traveling less on average, since many of them would not be going on the seasonal work trips for example. Nevertheless, since there is no data comparing the short term movement of adults and children in Senegal, I use an across the board weight of 1.4 to get an upper bound on movement.[15] The weighted data will provide an overestimate of total movement and an underestimate of the impact of each trip. I will provide both weighted and unweighted results in the analysis. While the data are based on SIM cards, and it is possible a SIM could represent several people or one

[15]This weight is used because the SIM cards are around 70% of the full population.

person might have multiple SIMs, I will use "individual" when referring to a SIM.[16]

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 enables tracking of people in space as they make calls from different tower locations. The data are 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. Due to the availability of only one year of cell phone data, the analysis is limited to the twelve months of 2013. Table 1 provides some statistics describing the phone data. On average there are 1657 calls or texts per person during the year, and on average a person has a call or text observation on 155 days.

Each mobile phone tower is assigned a health district based on its GPS coordinates. An individual is assigned a health district location on a given day based on the cell tower of the most recent call or text as of 11:59pm on the day in question.[17]. In instances where there are days with no calls, the health district location of the day closest to the one missing where a location is known based on a call/text is assigned to the missing days.[18] In this way a health district location is assigned to each SIM card for every day of the year.

Movement is defined as a change in location from one health district to another between two consecutive days. Table 1 also includes statistics on movement seen in

[16]As discussed earlier, while a phone may be used by multiple people, the majority of individuals each have their own SIM card, and so it is unlikely that a SIM represents multiple people. In addition, the survey finds that 9.6% of those with a phone cite using more than one mobile phone, and only 0.6% cite using more than two phones. So it is possible that someone might have multiple Orange SIM cards, but it would be a small percentage, since there is no benefit to having multiple Orange SIM cards unless they are used in different phones.

[17]This method was compared with another method of capturing individual locations used in the literature that uses 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 days for individuals when no calls were made in the time frame specified by the alternative method.

[18]Assignment can be based on a location before or after the missing day, depending on which day is closer.

the data. Movement is broken out into all movement, including between neighboring districts that share a border, and movement excluding trips between neighboring districts.The population is highly mobile, with over 80% of the population taking at least one trip and over a quarter million people traveling on average on any given day. This decreases down to 60% of the population taking at least one trip and almost 140,000 people traveling per day once trips between neighboring districts are removed. People on average take 10 different trips and visit almost five different health districts, though half of the trips are to neighboring districts and on average only three of the districts visited are non-neighboring. The large portion of trips between neighboring districts includes individuals that live on the border between two districts and might be making calls on either side of the border if locations they visit daily are located on both sides. While the main analysis is done using movement from all districts, I also run the analysis excluding trips between neighboring districts and provide the results in the robustness section.

Since the analysis is focused on the eight health districts for which there is malaria data at the health post level, I will only focus on movement into those eight districts. A move is still defined as a change in health district, so for each health district, the people entering are those that enter the district from any of the other 75 health districts. In addition to assigning the towers within these eight health districts to a health district, I also assign them to a healthpost catchment area based on their GPS location. Therefore, each traveler entering one of the eight health districts is assigned a specific health post catchment area that they enter based on their last call or text on the day they enter the district. If a person moves within a district to a different health post catchment area, this is not counted as a move.

Panel a of Figure 5 shows the average number of people entering one of the health post catchment areas each day as a percent of the population in that catchment area, along with vertical lines marking public holidays and important pilgrimages. The movement patterns largely align with the holidays and pilgrimages, showing that the majority of deviations in movement arise from special occasions that motivate indi-

viduals to travel, potentially to visit family and friends, or else to visit an important religious site. This pattern is in line with the findings in A. S. Fall, (1998) that the majority of migrants to Dakar visit their home area primarily for holidays, religious festivals and family ceremonies. On average for all the health post catchment areas, around 3 percent of the population enters on any given day. Turning to Panel b of the graph, the data are broken out by individual health post catchment areas for one district. This shows that the variation in percent of people entering can vary widely by health post catchment area and date. For health post catchment areas where an important religious leader resides, on certain religious holidays the number of people entering is close to or over 50% of the population of the area. For other health posts, the beginning of certain agricultural seasons or other holidays lead to large jumps in people entering. This variation makes it possible to study the impact of people entering on malaria cases in these areas that are otherwise geographically close together and very similar.

3 Model of Malaria

There is a large literature in epidemiology that models the spread of malaria, and I want to capture the basic nature of these models in my empirical specification. To understand the impact of travel on malaria incidence, it is necessary to break up cases into those generated locally and those imported from outside of the area. Additionally, some of the local cases will arise from mosquitoes that were infected by an imported case, which constitutes the negative externality. In order to separate out these different factors, I build on an existing epidemiological model. Basing my empirical specification on the model makes it possible to compare my coefficients to estimated parameters in the epidemiology literature.

Due to the two host system, malaria is modeled using two differential equations to describe the dynamics of infected humans and the dynamics of infected mosquitoes. These were first modeled by R. Ross, (1910) and then further expanded by Macdonald et al., (1957). The model used in this paper is a Ross-Macdonald type model

based on models used in D. L. Smith and McKenzie, (2004), Cosner et al., (2009) and Torres-Sorando and Rodriguez, (1997). There are two state variables and the model is usually expressed in continuous time, though here I will present it in discrete time. The first state variable is y_{it}, the fraction of mosquitoes infected in location i. The second state variable is the fraction of humans infected in location i, x_{it}.

In the model extension here that includes the impact of population movement, cases can either be generated in location i, L_i, or imported from any other location j into i, $\mathcal{I}_i$. Local infections in location i are generated based on the fraction of the population that is susceptible to malaria, S_i, the transmission efficiency from infected mosquitoes to humans, b, the number of bites on humans per mosquito per day, a_i, the ratio of mosquitoes to humans, m_i and the fraction of mosquitoes infected, y_i. w is the amount of time between a primary case and a secondary case generated by mosquitoes biting the primary person and eventually biting the secondary person. The change in the number of infected humans and mosquitoes is described by:

$$y_{it} - y_{it-w} = a_{it-w}cx_{it-w}(e^{-\mu_{it-w}\tau_{it-w}} - y_{it-w}) - \mu_{it-w}y_{it-w} \quad (1)$$

$$x_{it} - x_{it-w} = L_{it} + \mathcal{I}_{it} - rx_{it-w} = m_{it-w}a_{it-w}by_{it-w}S_{it-w} + \mathcal{I}_{it} - rx_{it-w} \quad (2)$$

where c is the transmission efficiency from infected humans to mosquitoes, μ_i is the mortality rate of mosquitoes and τ_i is the incubation period from the time a mosquito becomes infected until it is infectious. The parameter r is the recovery rate of humans.[19]

I make five main assumptions:

1. I assume the time necessary for a secondary case to appear is one month based on the description of the malaria cycle in section two

2. The mosquito population is at the steady state since mosquito populations

[19]Note that b, c and r are biological parameters that are assumed to not vary with time and location.

have a relatively rapid turnover [20]

3. All malaria cases in month t are treated immediately and infected individuals are completely removed within month t.[21]

4. The proportion of the population that is susceptible is one[22]

5. Based on D. L. Smith and McKenzie, (2004), when the proportion of infected humans is small, the the number of infectious bites received per day by a human (also known as the entomological inoculation rate, taking the form $EIR_i = \frac{m_i a_i^2 c e^{-\mu_i \tau_i} x_i}{\mu_i + a_i c x_i}$) can be approximated by $cC_i x_i$, where C_i is the expected number of humans infected per infected human per day, assuming perfect transmission efficiency (b=c=1), known as the vectorial capacity.

Based on assumption 1, I set $w = 1$. Based on assumption 2, I solve equation 1 for the quasi-equilibrium proportion of infectious mosquitoes as has been done in D. L. Smith and McKenzie, 2004 and Ruktanonchai et al., 2016:

$$y_{it-1} = \frac{a_{it-1} c x_{it-1} e^{-\mu_{it-1} \tau_{it-1}}}{\mu_{it-1} + a_{it-1} c x_{it-1}} \tag{3}$$

Assumption 3 implies the recovery rate, r, is equal to one since all individuals infected at time $t-1$ recover within month $t-1$. This allows me to focus on new cases of malaria in month t. Assumption 4 signifies that $S_{it-1} = 1$ Rewriting equation 2 to

[20] According to the CDC, adult female mosquitoes, which are the ones that bite humans and spread malaria, probably do not live more than 1-2 weeks in nature (Center for Disease Control, 2015).

[21] The incidence observed in the analysis is based on clinically diagnosed cases, which are provided with free antimalarial treatment upon diagnosis. The literature shows that by 24 days after treatment with even the slowest acting medication, all parasites are gone (Nosten and N. J. White, 2007, N. White, 1997).

[22] Since there is no actual immunity to malaria, anyone that is not currently infected is susceptible to the disease. In my data, the average population susceptible across all months and districts is 0.997. The appendix includes results where the empirical specification includes the actual susceptible population.

incorporate the implications of assumptions 1-4:

$$x_{it} - x_{it-1} = \frac{a_{it-1}^2 bcm_{it-1} e^{-\mu_{it-1}\tau_{it-1}} x_{it-1}}{\mu_{it-1} + a_{it-1} c x_{it-1}} + \mathcal{I}_{it} - x_{it-1} \tag{4}$$

Based on assumption 5, equation 4 can be rewritten as:

$$\begin{aligned} x_{it} &= L_{it} + \mathcal{I}_{it} \\ &= bEIR_{it-1} + \mathcal{I}_{it} \\ &= bcC_{it-1} x_{it-1} + \mathcal{I}_{it} \end{aligned} \tag{5}$$

Assumption 5 is important because by rewriting EIR_{it-1} as $cC_{it-1}x_{it-1}$, I explicitly incorporate the impact of the incidence last month, x_{it-1}, on the incidence in the current month using a linear functional form, which helps me approximately estimate the secondary cases generated this month by cases last month. The assumption can be used here because the analysis is conducted in a low malaria setting.[23] Based on equation 6, we also have that:

$$x_{it-1} = L_{it-1} + \mathcal{I}_{it-1} \tag{6}$$

Substituting equation 6 into equation 6 results in:

$$x_{it} = bcC_{it-1}(L_{it-1} + \mathcal{I}_{it-1}) + \mathcal{I}_{it} \tag{7}$$

This model enables me to separate the locally generated cases from imported cases. In addition, the locally generated cases can be further separated into those attributed to a mosquito infected by someone local and those attributed to a mosquito infected by an imported case of malaria. This breakdown allows for the calculation of both primary and secondary cases generated by travelers.

[23]The validity of the assumption will be discussed in more detail in the results section.

3.1 Empirical Model and Calibration

Now I outline the empirical specification derived from the model developed in the previous section that I will estimate using OLS. From the model, I want to estimate specifically bcC_{it-1}. Due to data limitations, since I want to causally estimate the secondary cases generated by an imported case, instead of estimating a bcC parameter for each location and each month, I estimate an average bcC parameter across locations and time. By estimating an average parameter, I am able to include health post area fixed effects, which allow me to control for unobservable characteristics of a health post area that might impact the number of malaria cases independently of imported cases. I can also include month dummies to control for the seasonality of malaria, as well as rainfall controls, in order to account for the fact that population movement and malaria can independently be related to rainfall and agricultural seasons. Therefore, using these controls allows me to identify the impact of imported cases based on deviations from the mean for a given area.

I use the mobile phone data to calculate expected imported malaria cases entering and divide them by the population of the area they enter to calculate the imported incidence. The likelihood of an infected case entering will depend on where an individual comes from, how long the person spent there and how long the person spends in the place entered. I make two assertions:

1. The likelihood a person is infected is based on the proportion of the month spent in the district, T_{jp}, and the monthly incidence rate, x_{jt}

2. The contribution of an imported case to a new location is calculated as a fraction of time spent in the location entered, T_{ip}

By using the incidence rate, x_{jt} (which includes both locally generated and imported malaria cases), rather than prevalence of the location, I account for the fact that a person p_t coming in at time t could either have been infected in the place she came from, j, or she could represent an imported case that came into j from somewhere else first before entering j and then i.

Using these two assumptions, my estimating equation is:

$$x_{it} = \beta_1 L_{it-1} + \beta_2 \mathbb{E}(\mathcal{I}_{it-1}) + \beta_3 \mathbb{E}(\mathcal{I}_{it}) + \alpha Z_{it} + \gamma_i + \delta_t + \epsilon_{it} \quad (8)$$

$$\mathbb{E}(\mathcal{I}_{it}) = \frac{1}{H_{it}} \sum_{j \neq i} \sum_{p_t \in j} T_{ip}(x_{jt} T_{jp}) \quad (9)$$

where γ_i are area fixed effects, δ_t are month dummies and ϵ_{it} represents idiosyncratic shocks. The matrix Z_{it} includes rainfall, rainfall lagged one month, and rainfall lagged two months, which capture both the agricultural seasons and changes in malaria incidence due to environmental factors. H_{it} is the population of location i in month t. I cluster errors at the health post catchment area level to account for the fact that errors are correlated within panels. I include a robustness check with both spatial and panel autocorrelated standard errors.

The main coefficients of interest are β_2 and β_3, which represent the number of secondary cases generated by infected travelers and the number of primary malaria cases imported by infected travelers. Based on the above model, we have that $\beta_3 = 1$, which can be tested directly. Additionally, based on the model, $\beta_1 = \beta_2 = bcC$. This relationship makes it possible to calibrate whether the magnitude of the coefficients estimated is reasonable using estimates of b, c and C drawn from the literature. Vercruysse, Jancloes, and Van de Velden, (1983) conduct a longitudinal survey to examine entomological and parasitological aspects of human malaria transmission in one area of Senegal. They find monthly estimates of the vectorial capacity over the course of a year. The average vectorial capacity, C, is 1.13. Based on Gething et al., (2011), average transmission from infected humans to mosquitoes, c, is 0.161. Finally, the transmission from mosquitoes to humans, b, is 0.05 based on the linear model in D. L. Smith, Drakeley, et al., 2010, which is closest to the model estimated in this paper. This gives an average daily value of 0.0091, or a monthly value of 0.273 for bcC, which I compare to the coefficients β_1 and β_2 in the analysis.

3.2 Calculation of Necessary Variables

x_{it} is calculated as number of cases per 1000 people. It is measured using the health post catchment area data on number of cases. Cases are divided by the monthly population of i. The monthly population is calculated by taking the annual health post area population, adding the number of people entering and subtracting the number of people leaving each month based on the mobile phone data.

The contribution of p_t to imported cases is calculated based on the incidence of the district the person is coming from and the length of time spent in district j and health post catchment area i. Only up to 15 days in the place the person comes from and up to 15 days in the place the person enters are considered since the incubation period is 15 days. This is done since if a person were bitten by an infected mosquito 15 days before entering, she would become symptomatic on the first day she enters, and if the person were bitten by an infected mosquito on the last day before she enters, the person would become symptomatic by the 15th day in the area entered. Therefore, T_{ip} is the proportion of 15 days an individual p spends in location i after entering i and T_{jp} is the proportion of the month person p spent in location j with monthly malaria incidence x_{jt} in month t. I use the detailed knowledge of the timing from the mobile phone data to factor in how many of the 15 days were in month t and how many in month $t-1$ and use the incidence both in month x_{jt} and x_{jt-1} to determine the probability the person was infected. Locally generated incidence, L_{it}, is calculated as the different between total incidence and imported incidence. I include a specification where both the imported cases and the total malaria cases are scaled by the region utilization rate, to account for the fact that utilization varies slightly in different parts of the country.

Due to the complicated nature of the imported incidence variable, I break down what explains the variation in this variable. Imported incidence combines information on travelers, the incidence where they are coming from, the timing in the place they enter and the place they leave, and the population in the place they enter. I

find that based on a partial R^2 of .33, the health post catchment area explains a large portion of the variation, while .43 of the variation is explained by the month. Jointly, they explain about half of the variation (R^2 of 0.56), implying that the other half of the variation in the variable of interest is coming from a combination of the month and the location receiving imported cases. This is important because it shows how the identification is coming from the unique combination of detailed data on travelers and the prevalence of where they are coming from.

3.3 Identification

For my identification to be correct, it is necessary that looking within a health post catchment area over time, any idiosyncratic shocks in malaria incidence are not correlated with expected imported or lag imported malaria incidence. As described in the section concerning population movement in Senegal, agricultural seasons and holidays are the two major reasons for travel. Agricultural seasons are strongly correlated with environmental factors, especially rainfall. In addition, rainfall could also affect the conditions for travel (quality of roads). As was shown in the graph of malaria cases and rainfall, there is also a very strong relationship between rainfall and malaria. Since I explicitly measure the relationship between rainfall and malaria incidence by including rainfall covariates in my specification, I am able to control for this potential confounder.

It is also possible that holidays, which increase population movement, could also affect malaria. For example, people might spend more time outside during holidays and expose themselves to mosquitoes. I will address this potential threat to identification using a placebo test. Specifically, I look to see whether I find an effect when I scale travelers by average monthly incidence in the country rather than by the incidence of where they came from.

There is the possibility of reverse causality. Increased incidence could make people more likely to travel in order to take care of relatives. Higher incidence might also

decrease travel if people avoid going to areas with higher malaria. This will be studied by looking at the relationship between future imported cases and current malaria.

Finally, since I am including a lag of the dependent variable within my specification, it is possible that this could be correlated with the error term if there is an autocorrelated structure to the data, which would bias my estimates. I specifically test for autocorrelation using a Cumby-Huizinga general test and fail to reject that there is no autocorrelation.

4 Results

4.1 Quantifying Effect of Imported Cases

Table 2 shows the results from running the main specification. Column 1 of the table shows the specification looking at the five low malaria districts, where imported cases have been weighted to represent the full population. In these districts, there is a significant effect of travelers both on primary and secondary cases. For each imported case of malaria, there is around 1.161 cases reported at a health post. In addition, for each imported case, there is an additional 0.437 of a case generated the following month, representing the negative externality. The table also shows that for each local case, there is 0.347 of a case generated the following month as well.

In Column 2 of Table 2, I focus on the three high malaria districts. For these districts, although the sum of the coefficients on imported and lagged imported cases is not statistically different from the sum of the coefficients for the low districts, neither is significant. This is due to the very large standard errors for these variables. One explanation for this is that assumption 5 in the model section is violated. This assumption requires that the incidence is low in order for the vectorial capacity times transmission and incidence to approximate the entomological inoculation rate. This can be directly checked by estimating the entomological inoculation rate, EIR and the vectorial capacity times transmission and incidence, cCx for each health post

area in each month.[24] Figure 6 shows that for the low malaria districts, the EIR and cCx are very similar and lie along the 45 degree line. For the high malaria districts, as malaria incidence gets larger, the cCx grows at a faster rate than EIR, and the approximation no longer holds. Since the necessary assumptions do not all hold for the high malaria districts, the results are inconclusive. For local cases, the estimate is even larger for secondary cases generated (and significantly different from the coefficient on low malaria districts at the .005 level). Since in the high malaria districts there are more mosquitoes present that can transmit the disease, the larger coefficient, signifying higher secondary transmission, is in line with expectations.

Column 3 of Table 2 is focused on the low districts, but this time it scales both imported incidence and total incidence by health post utilization at the region level. The results are similar and not significantly different from the results in column 1. As discussed, utilization is relatively uniform throughout the country; therefore, the same results are captured whether or not utilization is incorporated. Finally, Column 4 uses the imported cases without scaling by health post utilization and without weighting the mobility data to be representative of the full population. This would make the assumption that the only movement in the country is the movement I see in the data, which would be an underestimate. As discussed, weighting the data assumes that the population without mobile phones and Sonatel SIM cards has similar movement patterns to the population captured in the data. Since the majority of the non-represented population is children that likely travel less than the adults, this represents an upper bound for the level of movement. The actual movement that occurs, and therefore, imported incidence, is then somewhere in between these lower and upper bounds. Thus the coefficients should also lie between these upper and lower bounds of between 1.2 and 1.7 for the primary effect and 0.44 and 0.63 for the negative externality. Appendix A contains regression results when lagged local and imported incidence is adjusted by the proportion of the population susceptible.

[24]These were estimated using parameter values from the literature for the various biological malaria parameters

Figure 7 shows the results averaged across health posts by district and split into low malaria and high malaria districts. Panels a and c look at how well the model fits the actual data, comparing average actual incidence in 2013 to incidence predicted by the model. Predicted incidence is calculated based on 500 draws of the parameters from their respective distributions.[25] I used the mean predicted incidence across all draws. Actual incidence is shown in the dark green line. Panels a and c show that the model provides a good approximation for average actual malaria incidence for both the low and high malaria districts.

Panels b and d of Figure 7 demonstrate how large the effect of imported incidence is on average in each district. This is done by comparing the predicted incidence from panels a and b to the incidence that would be predicted if there were no imported cases of malaria. Using the same parameters from the 500 draws, this predicted incidence with no imported cases was calculated by setting imported and lagged imported incidence to 0. For the low malaria districts, especially Richard Toll and Saint-Louis, imported incidence represents a substantial portion of the malaria incidence. On average annually per health post, travel represents 34% of the incidence. In the high malaria districts, while the absolute size of the effect of imported incidence is similar to the low malaria districts, it makes up a very small amount of the overall incidence in those districts. On average annually per health post, imported incidence only makes up around 3% of cases in the high malaria districts.

The effect of imported incidence as a percent of total incidence is much larger in the low malaria as compared to the high malaria districts. Since the goal of the government is to bring the low malaria districts down to complete elimination, targeting travelers needs to be an important component of this strategy since they represent around one third of cases in these districts. On the other hand, in the high malaria districts where the government wants to control malaria and decrease it significantly before it is ready to undertake the targeted policies used for elimination, targeting

[25]The distributions of the parameters were calculated using 10,000 bootstrap simulations of the model.

travelers would not be an effective strategy. In these districts, imported cases are only a small part of the problem and it is necessary to focus on larger, more widespread interventions first. Therefore, for the remaining portion of the paper, I will focus on the five low malaria districts where the government wants to implement targeted elimination strategies.

4.2 Testing the Model and Case Study Analysis

Focusing on the low malaria districts, Table 2, Panel B shows results from hypothesis tests corresponding to the expectations from the calibration in the model section. I first test whether the coefficient on imported incidence is significantly different from 1, since the model implies that each imported case per capita should contribute a case per capita to the incidence in the area entered. With a p-value of 0.78 on the Wald test, the coefficient is not significantly different from 0. I also test whether the coefficients on lagged imported incidence and lagged local incidence are each not significantly different from the epidemiological parameter prediction of 0.273, and I find that they are not. Finally, I test for equality between the coefficients on lag imported and lag local incidence and find that they are not significantly different from each other, which was the last prediction based on the model.

In addition to using estimated epidemiological parameters to compare to the size of the coefficients, it is also possible to use a case study in the district of Richard Toll in Senegal, for which data was collected on all primary and secondary cases of malaria starting in 2013.[26] This study can provide some measure of the secondary cases generated by a primary case, which can be compared to the magnitudes I find in my analysis. Table 3 shows the case numbers for each year between 2013 and 2015, split between imported and local as well as the number of secondary cases

[26]Details of this data collection are available in Appendix B. This type of data collection is very costly in both time and resources because it requires sending out health workers to individual households and tracking down the people that live there and in the proximity. The analysis in this one district can help validate how well the mobile phone data analysis does. Nevertheless, doing this type of survey on a large scale would be much more costly as compared to using the mobile phone data over a larger geographic area.

found through the investigations of the primary cases. The table shows that for each primary imported case there are on average between 0.09 and 0.15 secondary cases and for each local case there are on average between .12 and .23 secondary cases. These values are lower than what the main regression shows for secondary cases generated in the following month. Nevertheless, the value for imported cases is within the 95 percent confidence interval for imported cases in my analysis, and the upper limit for local cases is within the confidence interval for local cases. Appendix B further explains how policies implemented in Richard Toll can help explain the lower values seen there.

4.3 Mechanism Evidence

An issue with interpreting these results causally is that there could be reverse causality, whereby malaria incidence increases population movement. This could happen if people travel to take care of sick relatives for example. Figure 8 shows that malaria prevalence is not associated with future travel though. The graph shows coefficients from a regression of malaria prevalence on imported incidence going back two periods and going forward two periods, focused on low malaria districts.[27] Malaria incidence two months earlier and one month earlier has no relationship with imported cases in the current period. Malaria incidence in the current period and one month later is associated with higher malaria prevalence, as both of those coefficients jump. Imported cases do not seem to have a significant impact on incidence two months later, though the standard error on the coefficient is large. Since the sample size is much smaller after including the leads and lags, the standard errors are much larger, but the trend in the size of the coefficient still demonstrates what would be expected, with no impact of imported cases until they actually enter and secondary impact in the following month only.

I conduct several placebo tests to refute potential alternative explanations. Table 4

[27] Additional leads and lags could not be included because the sample size becomes too small since only twelve months of data are available.

shows the results from these. When using alternative measures of imported cases, I focus on imported cases in the current month and control for lagged total incidence.[28] The first column shows the actual specification used with imported cases calculated based on the incidence of where travelers enter from and the lengths of time they spend in the location they enter from and in the location they enter.

One alternative explanation is that the imported incidence variable is just picking up increased travel during certain times in certain locations, such as religious holidays and locations with important religious leaders, and those times might be correlated with higher malaria for some alternative reason. For example, during these religious holidays people might spend more time outside and are more likely to be bit by mosquitoes. To test this, rather than using the incidence of the location a person is coming from to calculate their probability of importing malaria, I instead use the average monthly incidence across all health districts. In this way, the location of where travelers enter from no longer affects the variable, only the travel patterns do. When I do this, there is no relationship between this alternative variable and malaria incidence, as shown in Column 2 of Table 4.

The flip side to the explanation that only the travel matters is that only the incidence of where people are entering from matters. This could be the case if for example travelers tend to enter from locations that have similar incidence to the place they enter, and therefore the variable is capturing this relationship between the locations coming from and going to, and is not related to the travel itself. To test this, I calculate expected imported incidence based on the incidence of the locations where people are entering from. Rather than separately calculating a probability for each person using the length of travel, I use an average value for time spent in place of origin and time spent in destination, and assign an average number of trav-

[28] I do not separate local and imported lagged incidence because the lagged local is calculated as the remainder from the lagged total less lagged imported, and in the placebo tests, the measure for imported is at times several times greater than total cases;therefore, picking out a local effect is not possible.

elers to be going from the origin to destination.[29] Column 3 of Table 4 shows that when imported cases are calculated for an average number of travelers per destination/origin pair, there is no longer a relationship between this variable and malaria incidence. These two tests demonstrate that it is the interaction of location, number of travelers and time spent that is necessary to produce the higher malaria incidence.

In Columns 4 and 5, I test that the effect is only seen for malaria incidence. In Column 4, I use a variable for imported disease incidence other than malaria. Other disease incidence is calculated based on number of people coming to a health post for a consultation, removing any cases of fever that could be symptomatic of malaria, and scaled by the population of the health post area. Instead of using malaria incidence to calculate the contribution to an infection, I use this disease incidence to calculate the contribution to a non-malarial disease. Column 4 shows that this variable has no relationship with malaria incidence in the location where diseases are being imported. In Column 5, I use the original imported malaria incidence variable as the regressor, but I change the dependent variable to be disease incidence excluding malaria, instead of malaria incidence, and I use lagged disease incidence instead of lagged malaria incidence. Imported malaria incidence does not seem to have a relationship with overall disease incidence (excluding malaria). Since the regressors and dependent variables in the different placebos are on different scales, I also include the beta coefficients, which are measured in terms of standard deviations and make it possible to compare coefficients across all five regressions. The beta coefficients are five to sixty times smaller for the placebo test imported variables as compared to the base specification.

[29] Average number of travelers was calculated based on total number of travelers per month and number of unique origin destination pairs in that month, and the length of trips was also calculated based on the average length across all travelers in a month.

5 Policy Targeting Strategies

I now turn to which travelers would be cost effective to target in order to reduce incidence in the five low malaria districts. Suppose an average cost C per traveler targeted is assigned. In this framework, if there were no targeting and all travelers to the 5 pre-elimination districts were treated, then 6,956,197 travelers would need to be targeted and 602 malaria cases, representing around 40% of total cases in this area, would be treated or prevented. Even if the cost were only $0.50 per traveler (slightly lower than the cost of an RDT), this would amount to close to $6,000 per case treated or averted. The ProACT policy discussed in the background section, which sent out health workers weekly to test potential cases, cost a total of $8000 for the pilot and using back of the envelope calculations was around $38 per case of malaria averted (*Senegal Malaria Operational Plan FY 2015* 2015, Linn et al., 2015). This program was in an area with relatively high malaria, therefore the number of cases averted is much higher than if it were to be done in a very low malaria area like the five health districts studied in the current analysis. Nevertheless, it demonstrates that a policy directed at all travelers would have a cost by far exceeding other potential interventions such as ProACT.

Therefore, if a policy were to target travelers, only a subset of the most cost effective travelers should be targeted. I make the simplifying assumption that targeting of travelers would occur based on the district the individuals are coming from and the month in which they are traveling and entering a pre-elimination area. Therefore, if a district-month is chosen as receiving targeting, then every person traveling from that district to the pre-elimination area in that month would be treated with the policy. This type of targeting could be implemented with the SMS policy that relies on the mobile operator to send alerts to people when they enter a certain zone depending on where they are entering from or a policy of health workers meeting buses from certain locations at the bus terminal upon arrival. While the particular details of a specific policy will not be discussed here, the targeting analysis can shed light on what types of policies might work and be cost effective.

Figure 9 portrays the relationship between number of people targeted and number of cases averted for four different targeting strategies. Cases averted is calculated using the results from the previous section.[30] For all strategies, I make the assumptions that whatever policy is chosen to target, there is a uniform cost per person that is the same for all strategies and the policy works with 100% effectiveness across all strategies. For the first strategy, shown in green on the graph, the government randomly chooses district-months that would be targeted. The second targeting strategy I study, shown in orange, is one where the government chooses district-months randomly within the malarial season (August to December) before randomly choosing district-months during the rest of the year.[31] The third strategy considered, in gray on the graph, is one where the government uses monthly malaria incidence per thousand people in the prior year to order the districts from those with the highest to the lowest incidence. Finally, the fourth strategy uses the analysis in the previous section to calculate a cost-effectiveness value for each district-month. To calculate cost-effectiveness, I make the simplifying assumption that the cost of targeting an individual traveler is the same no matter where they come from. Therefore, the cost-effectiveness is how many travelers are coming in from a particular district-month and would need to be targeted (the cost) and how many cases of malaria would be treated and averted if these travelers were targeted (the benefit).

The top panel of Figure 9 shows the full cost curve, from a strategy where only one district-month is treated, all the way up to all district-months being treated. Depending on the resources available to the government and the cost of targeting one traveler, it is possible to determine for any given budget how many cases would be treated or averted depending on the type of targeting strategy chosen and which district-months are included. Zooming in on the curve below 100,000 travelers, the strategies of targeting randomly or randomly during the malarial season are ex-

[30] I run 1000 simulations based on parameter values drawn from their distribution and calculate the total primary and secondary imported cases expected from each district-month.

[31] Appendix figure A.1 shows the large variance in number of people treated per case averted from month to month.

tremely inefficient. Randomly targeting districts during the malaria season is the closest strategy for travelers that was attempted in the district of Richard Toll using volunteers to alert health workers of new travelers[32]. In this simulation, I find that with this random targeting during the malarial season, if around 7000 travelers are targeted, we would expect around 3 cases of malaria averted. This is somewhat similar to the results in Richard Toll where around 3,500 travelers were targeted with testing and 10 of them were positive cases. If instead, 3,500 travelers were targeted based on the cost-effective strategy, around 50 cases of malaria would have been treated or averted, which would be a five-fold increase.

How much better the cost-effective strategy does as compared to the strategy utilizing information on past incidence depends on the number of travelers targeted. As Panel b of Figure 9 shows, if around 50,000 travelers are targeted, the benefit using the cost-effective policy averts around 30% more cases as compared to the next best policy of using the information on incidence from the previous year and almost 8 times more cases are averted as compared to randomly targeting during the malarial season. While not as good as the cost-effective strategy, the simulation shows that the next-best strategy of using the existing malaria data is also effective at decreasing the case load.

6 Robustness Checks

I do several robustness checks to test the main specification in Table 5. First, I rerun the analysis using cases rather than incidence. The model remains the same, with the only change being that both sides are multiplied times the population in location i. Therefore, the coefficients should remain around the same size since the hypothesis for β_1, β_2 and β_3 are the same. Column 1 shows the baseline specification while Column 2 shows the specification using cases rather than incidence that shows that, indeed, the coefficients are very similar to the main specification using

[32]This strategy is described in more detail in the background section

incidence. β_3 is still not significantly different from 1, although it is now significant at the .01 level because the standard error has decreased and the coefficient is larger. The coefficient on lagged local cases is almost identical to the main specification, and while the coefficient on lagged imported cases is slightly smaller, neither is significantly different from the epidemiological prediction of 0.273, and the two coefficients are still not significantly different from each other (Wald test has p value of 0.645).

In Column 3, I rerun the specification using Conley standard errors that account for spatial autocorrelation and serial autocorrelation over time (Hsiang, 2010).[33] In the results, it looks like the standard error on current imported incidence becomes smaller, while the standard error for lagged imported incidence goes up. When correcting for the spatial and autocorrelation, though, it is not possible to account for panel clustering as well. Clustering at the health post catchment level, used throughout the analysis, actually decreases the standard errors for lagged imported incidence. Therefore, the larger standard error in column two for lagged incidence is actually due to the lack of clustering rather than the spatial or panel autocorrelation.

In Column 4, rather than using time varying population based on mobile phone subscribers going in and out of a health post area, I instead only use the static census population. This makes very little difference, as the coefficients are not statistically different from the ones in the baseline model. In Column 5, I use net imported incidence, where I subtract potentially infected travelers leaving the health post catchment area. This helps to remove some of the noise from individuals that might be living on the border between two districts and traveling back and forth constantly, since on net, their travel would be close to 0. Both the coefficients on imported and lagged imported do not change significantly and they remain significant. Finally, in Column 6, I more explicitly try to account for individuals that might be living on the border between two districts and traveling constantly between the two (as was shown in the data section, travel between neighboring districts makes up almost half of all travel). Since malaria is not determined by borders but by

[33] I use a 30km cutoff for the spatial correlation and two lags for the autocorrelation.

geographic location, the prevalence close to the border on either side should not be very different, and therefore, travelers on the border that go back and forth within a small radius are not expected to impact the incidence. Column 6 shows that the coefficients remain significant but increase and the standard errors get larger on both current and lagged imported incidence. This could indicate that neighboring travelers are important for malaria transmission, and omitting them leads to an upward bias in the coefficient.

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 in order to study how policies could be targeted effectively to mitigate this negative externality. 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 incidence. 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 are 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 eliminating 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 for each imported case of malaria, there are around 1.6 cases of malaria that appear at a healthpost in a pre-elimination area. These findings are used to study the cost effectiveness of four different targeting strategies for travelers. I find that targeting the most cost effective districts during certain months results in the largest drop in cases, with up to five times as many cases averted as compared to the current strategy aimed at travelers. This is the first study to evaluate the cost

effectiveness of different targeting schemes directed at travelers.

While the work here focuses on malaria, it is possible to implement these types of models for other infectious diseases such as influenza, Ebola, or Zika. 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 policies, 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 are already being collected. This could help countries more effectively fight existing infectious diseases and prevent epidemics of new diseases.

References

Adda, Jérôme (2016). "Economic Activity and the Spread of Viral Diseases: Evidence from High Frequency Data". In: *The Quarterly Journal of Economics* 131.2 (2016), pp. 891–941.

Adriansen, Hanne Kirstine (2008). "Understanding pastoral mobility: the case of Senegalese Fulani". In: *The Geographical Journal* 174.3 (2008), pp. 207–222.

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. 2014.

Alegana, Victor A et al. (2013). "Estimation of malaria incidence in northern Namibia in 2009 using Bayesian conditional-autoregressive spatial–temporal models". In: *Spatial and spatio-temporal epidemiology* 7 (2013), pp. 25–36.

Alkhalife, Ibrahim S (2003). "Imported malaria infections diagnosed at the Malaria Referral Laboratory in Riyadh, Saudi Arabia." In: *Saudi medical journal* 24.10 (2003), pp. 1068–1072.

ANSD and ICF International (2015). *Senegal DHS, 2014 - Final Report Continuous 2012-14*. Tech. rep. 2015.

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 (2009), pp. 21484–21489.

Bleakley, Hoyt (2007). "Disease and development: evidence from hookworm eradication in the American South". In: *The Quarterly Journal of Economics* 122.1 (2007), p. 73.

— (2010). "Malaria eradication in the Americas: A retrospective analysis of childhood exposure". In: *American Economic Journal: Applied Economics* 2.2 (2010), pp. 1–45.

Bogoch, Isaac I et al. (2016). "Anticipating the international spread of Zika virus from Brazil." In: *Lancet (London, England)* 387.10016 (2016), pp. 335–336.

Bousema, Teun et al. (2012). "Hitting hotspots: spatial targeting of malaria for control and elimination". In: *PLoS Med* 9.1 (2012), e1001165.

Canning, David (2006). "The economics of HIV/AIDS in low-income countries: the case for prevention". In: *The Journal Of Economic Perspectives* 20.3 (2006), pp. 121–142.

Casey, Nicholas (2016). "Hard Times in Venezuela Breed Malaria as Desperate Flock to Mines". In: *The New York Times* (Aug. 2016). URL: `http://www.nytimes.com/2016/08/15/world/venezuela-malaria-mines.html?smid=nytcore-ipad-share%5C&smprod=nytcore-ipad%5C&%5C_r=0`.

Center for Disease Control (2015). *Anopheles Mosquitoes*. http://www.cdc.gov/malaria/about/biology/ Accessed: 2015-05-06. 2015.

Center, Climate Prediction (2016). *Climate Prediction Center (CPC) Rainfall Estimator (RFE) for Africa*. ftp://ftp.cpc.ncep.noaa.gov/fews/fewsdata/africa/rfe2/geotiff/. 2016.

Cho, Eunjoon, Seth A Myers, and Jure Leskovec (2011). "Friendship and mobility: user movement in location-based social networks". In: *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM. 2011, pp. 1082–1090.

Cohen, Justin M, Sabelo Dlamini, et al. (2013). "Rapid case-based mapping of seasonal malaria transmission risk for strategic elimination planning in Swaziland". In: *Malaria journal* 12.1 (2013), p. 1.

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

Cosner, C et al. (2009). "The effects of human movement on the persistence of vector-borne diseases". In: *Journal of theoretical biology* 258.4 (2009), pp. 550–560.

Deshingkar, Priya and Sven Grimm (2005). *Internal migration and development: A global perspective*. 19. United Nations Publications, 2005.

Djibo, Yacine and Fara Ndiaye (2013). *Teaming up Against Malaria*. Tech. rep. 2013.

Doolan, Denise L, Carlota Dobaño, and J Kevin Baird (2009). "Acquired immunity to malaria". In: *Clinical microbiology reviews* 22.1 (2009), 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, 2013.

Epstein, Joshua M et al. (2007). "Controlling pandemic flu: the value of international air travel restrictions". In: *PloS one* 2.5 (2007), e401.

Fall, Abdou Salam (1998). "Migrants' long-distance relationships and social networks in Dakar". In: *Environment and Urbanization* 10.1 (1998), pp. 135–146.

Fall, Papa Demba, María Hernández Carretero, and Mame Yassine Sarr (2010). "Country and Research Areas Report". In: (2010).

Franckel, Aurélien and Richard Lalou (2009). "Health-seeking behaviour for childhood malaria: household dynamics in rural Senegal". In: *Journal of biosocial science* 41.01 (2009), pp. 1–19.

Gething, Peter W et al. (2011). "A new world malaria map: Plasmodium falciparum endemicity in 2010". In: *Malar J* 10.378 (2011), pp. 1475–2875.

Goldsmith, Peter D, Kisan Gunjal, and Barnabe Ndarishikanye (2004). "Rural–urban migration and agricultural productivity: the case of Senegal". In: *Agricultural economics* 31.1 (2004), pp. 33–45.

Herrera, Catalina and David E Sahn (2013). "Determinants of internal migration among Senegalese youth". In: *Cornell Food and Nutrition Policy Program Working Paper* 245 (2013).

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

Hsiang, Solomon M (2010). "Temperatures and cyclones strongly associated with economic production in the Caribbean and Central America". In: *Proceedings of the National Academy of sciences* 107.35 (2010), pp. 15367–15372.

Huang, Zhuojie, Andrew J Tatem, et al. (2013). "Global malaria connectivity through air travel". In: *Malar J* 12.1 (2013), 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 (2013), 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 (2006), pp. 38–45.

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

Lépine, Aurélia and Alexis Le Nestour (2012). "The determinants of health care utilisation in rural Senegal". In: *Journal of African economies* (2012), ejs020.

Lima, A et al. (2015). "Disease containment strategies based on mobility and information dissemination". In: *Scientific reports* 5 (2015), p. 10650.

Linares, Olga F (2003). "Going to the city...and coming back? Turnaround migration among the Jola of Senegal". In: *Africa* 73.01 (2003), pp. 113–132.

Linn, Annē M et al. (2015). "Reduction in symptomatic malaria prevalence through proactive community treatment in rural Senegal". In: *Tropical Medicine & International Health* 20.11 (2015), pp. 1438–1446.

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

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

Lynch, Caroline A et al. (2015). "Association between recent internal travel and malaria in Ugandan highland and highland fringe areas". In: *Tropical Medicine & International Health* 20.6 (2015), pp. 773–780.

Macdonald, George et al. (1957). "The epidemiology and control of malaria." In: *The Epidemiology and Control of Malaria.* (1957).

Ndiath, Mamadou O et al. (2012). "Low and seasonal malaria transmission in the middle Senegal River basin: identification and characteristics of Anopheles vectors". In: *Parasites & vectors* 5.1 (2012), p. 1.

Nosten, François and Nicholas J White (2007). "Artemisinin-based combination treatment of falciparum malaria". In: *The American journal of tropical medicine and hygiene* 77.6 Suppl (2007), pp. 181–192.

Okell, Lucy C et al. (2012). "Factors determining the occurrence of submicroscopic malaria infections and their relevance for control". In: *Nature communications* 3 (2012), p. 1237.

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 (2004), 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 (2012), pp. 1025–1058.

Pison, Gilles et al. (1993). "Seasonal migration: a risk factor for HIV infection in rural Senegal". In: *JAIDS Journal of Acquired Immune Deficiency Syndromes* 6.2 (1993), pp. 196–200.

PNLP, INFORM and LSHTM (2015). *Senegal: A Profile of Malaria Control and Epidemiology*. Tech. rep. 2015.

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

Ross, Ronald (1910). *The prevention of malaria*. Dutton, 1910.

Ruktanonchai, Nick W et al. (2016). "Identifying malaria transmission foci for elimination using human mobility data". In: *PLoS Comput Biol* 12.4 (2016), e1004846.

Russell, Paul F and Domingo Santiago (1934). "Flight range of the funestus-minimus subgroup of anopheles in the Philippines". In: *The American Journal of Tropical Medicine* 14.2 (1934).

Senegal Malaria Operational Plan FY 2015 (2015). Tech. rep. President's Malaria Initiative, 2015.

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 (2010), pp. 391–397.

Smith, David L, Chris J Drakeley, et al. (2010). "A quantitative analysis of transmission efficiency versus intensity for malaria". In: *Nature communications* 1 (2010), p. 108.

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

Stuckler, David et al. (2011). "Mining and risk of tuberculosis in sub-Saharan Africa". In: *American journal of public health* 101.3 (2011), pp. 524–530.

Tam, Clarence C, Mishal S Khan, and Helena Legido-Quigley (2016). "Where economics and epidemics collide: migrant workers and emerging infections". In: *The Lancet* (2016).

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 (2009), 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 (2010), pp. 12222–12227.

The Global Fund (2016). *Financials*. 2016. URL: `http://www.theglobalfund.org/en/financials/`.

Thomas, Christopher J, Dónall E Cross, and Claus Bøgh (2013). "Landscape movements of anopheles gambiae malaria vector mosquitoes in Rural Gambia". In: *PloS one* 8.7 (2013), e68679.

Torres-Sorando, Lourdes and Diego J Rodriguez (1997). "Models of spatio-temporal dynamics in malaria". In: *Ecological modelling* 104.2 (1997), pp. 231–240.

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

Vercruysse, J, M Jancloes, and L Van de Velden (1983). "Epidemiology of seasonal falciparum malaria in an urban area of Senegal". In: *Bulletin of the World Health Organization* 61.5 (1983), p. 821.

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

Wesolowski, Amy, CJE Metcalf, et al. (2015). "Quantifying seasonal population fluxes driving rubella transmission dynamics using mobile phone data". In: *Proceedings of the National Academy of Sciences* 112.35 (2015), pp. 11114–11119.

Wesolowski, Amy, Taimur Qureshi, et al. (2015). "Impact of human mobility on the emergence of dengue epidemics in Pakistan". In: *Proceedings of the National Academy of Sciences* 112.38 (2015), pp. 11887–11892.

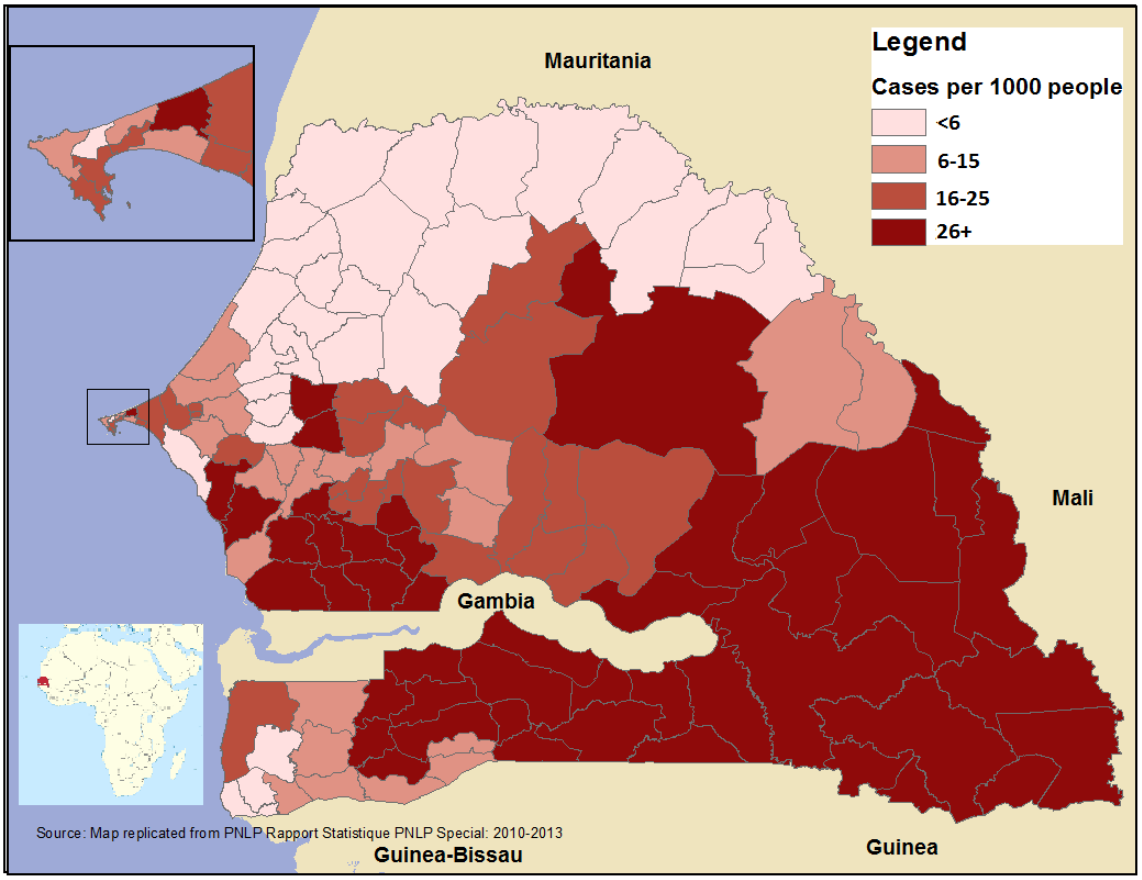
White, NJ (1997). "Assessment of the pharmacodynamic properties of antimalarial drugs in vivo." In: *Antimicrobial agents and chemotherapy* 41.7 (1997), p. 1413.

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

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

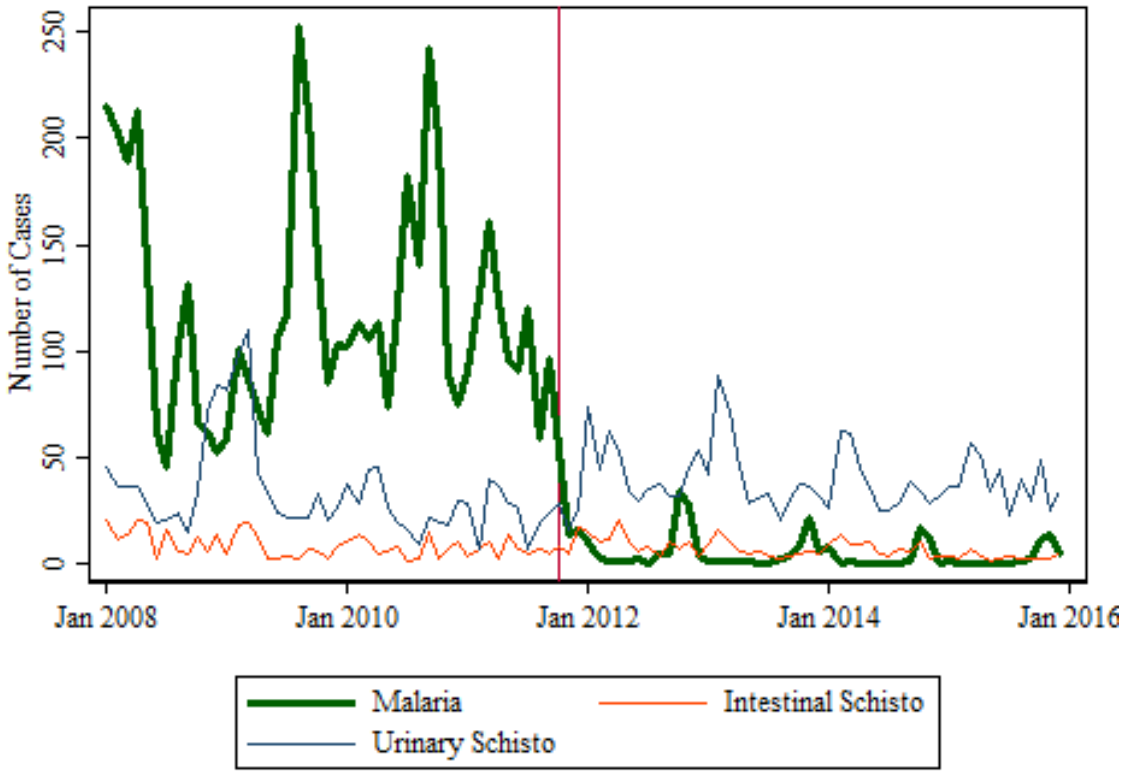
Yukich, Joshua O et al. (2013). “Travel history and malaria infection risk in a low-transmission setting in Ethiopia: a case control study”. In: *Malar J* 12 (2013), p. 33.

Figure 1: Map of Malaria Incidence in Senegal in 2013



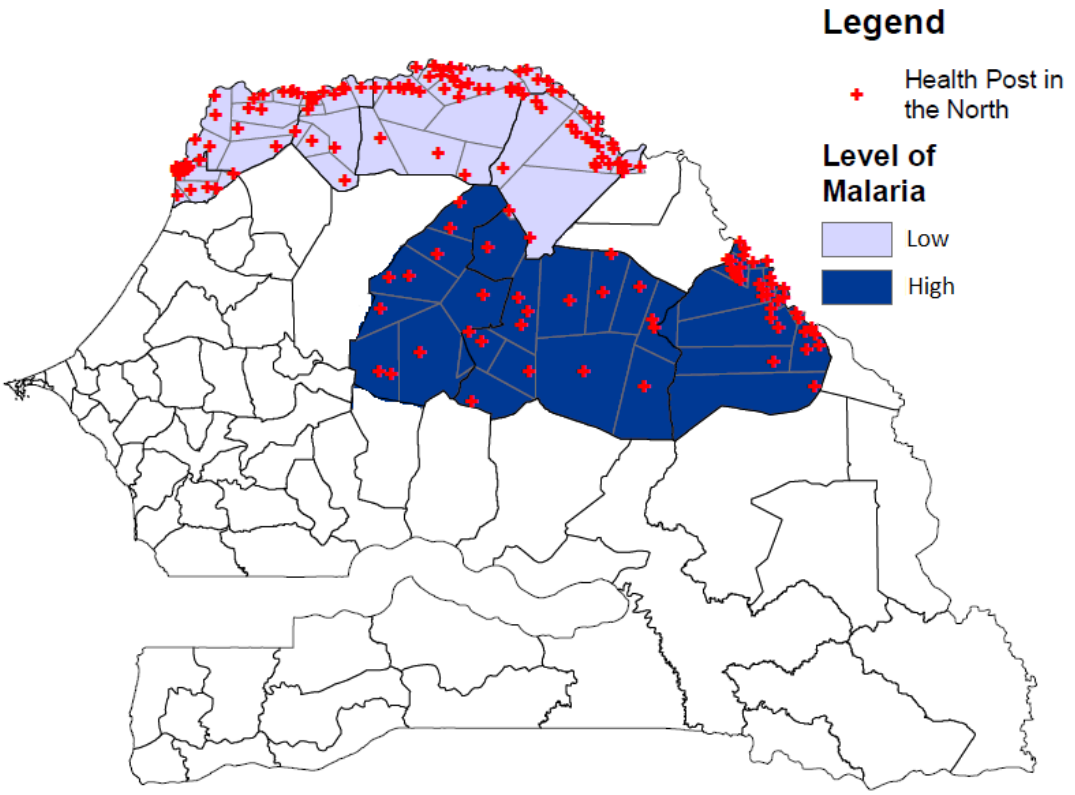
Notes: Data comes from tested and confirmed cases at the health district level compiled by the PNLP.

Figure 2: Effect on Malaria Cases of a Policy Targeting Migrant Workers at the Senegalese Sugar Company



Notes: The figure shows number of cases of malaria and two types of schistosomiases seen at the health post of the Senegalese Sugar Company. The red vertical line marks the timing of when a new policy was implemented by the company that tested every migrant worker and treated those that were infected.

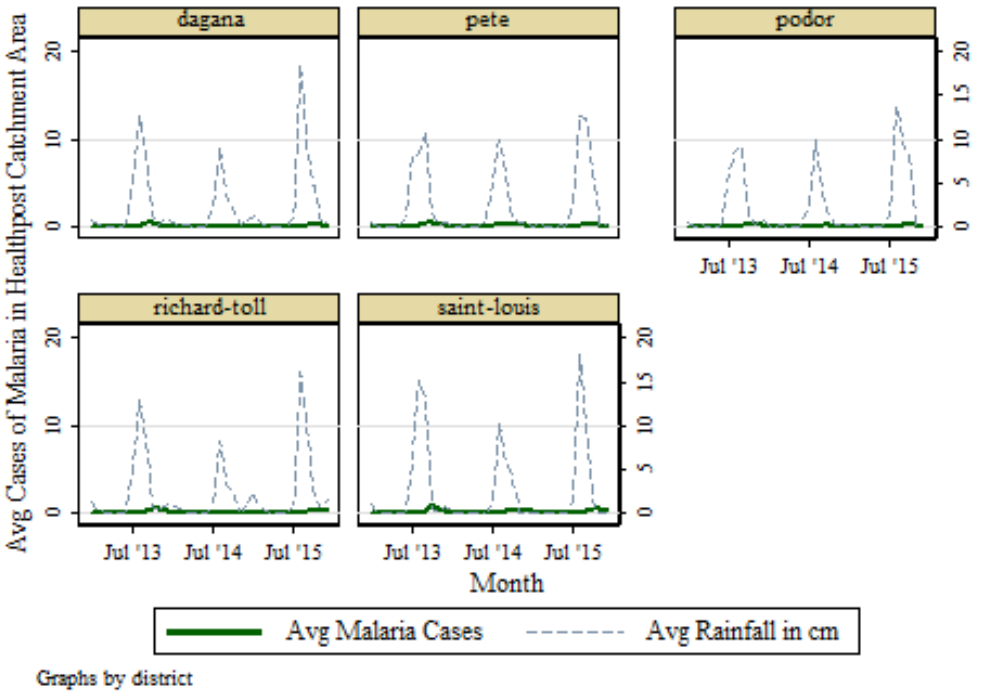
Figure 3: Senegal Health Districts and Location of Health Posts in the North for Which There is Monthly Data



Notes: The eight health districts with data available at the health post level are subdivided into health post catchment areas that group health posts and mobile phone towers together.

Figure 4: Average Monthly Malaria Cases per 1000 in Healthpost Catchment Area and Average Monthly Rainfall, Sept 2012-Dec 2015 by Health District

(a) Low Malaria Districts



(b) High Malaria Districts

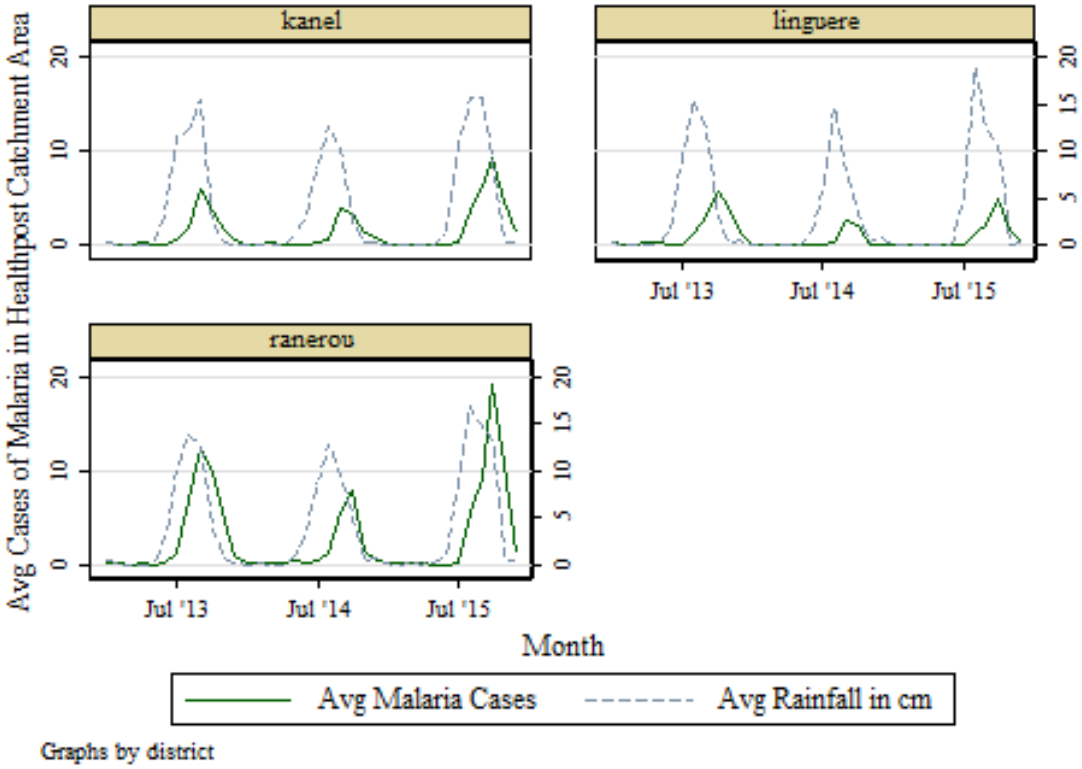
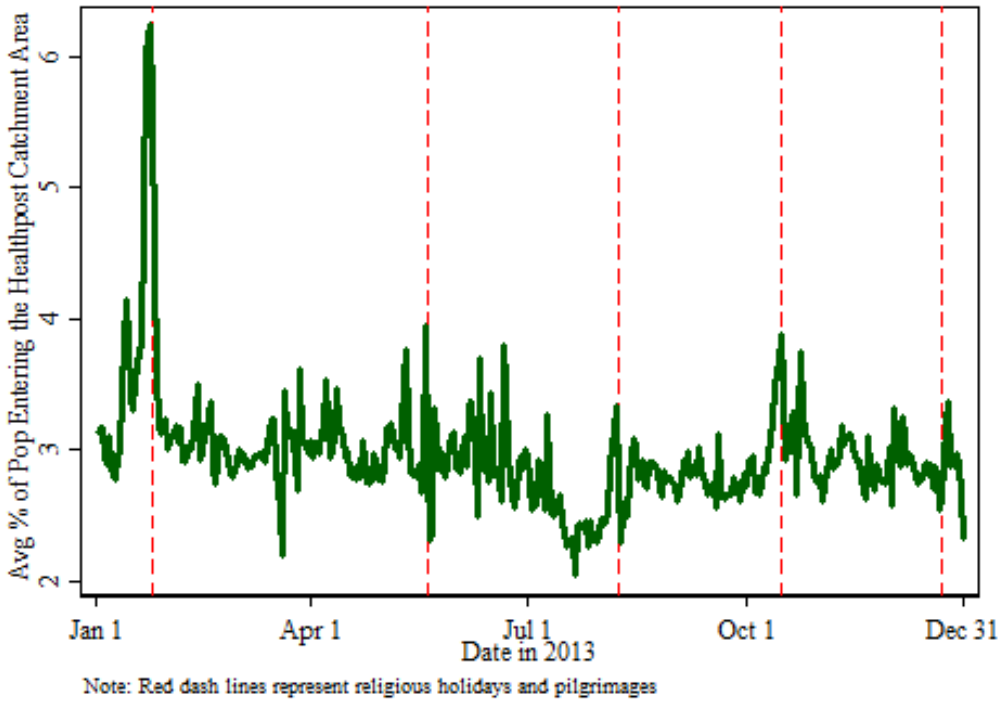


Figure 5: Percent of People Entering a Health Post Catchment Area

(a) Avg Across Health Post Catchment Area



Note: Red dash lines represent religious holidays and pilgrimages

(b) Healthpost Catchment Areas in Richard Toll

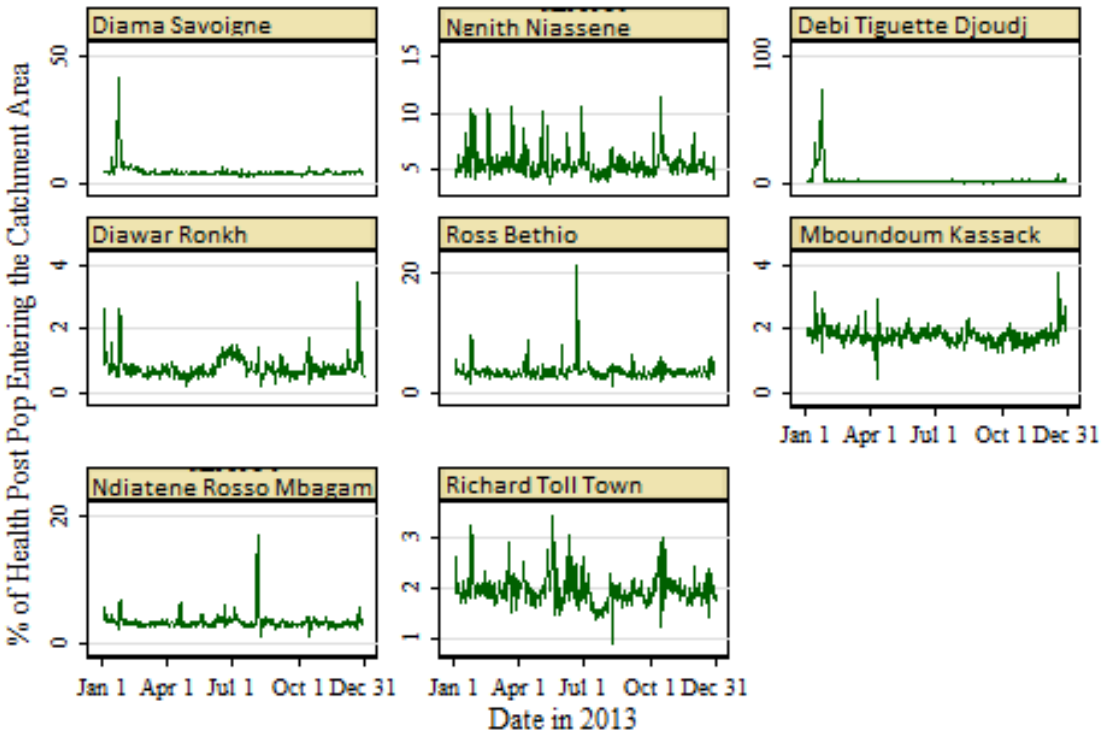
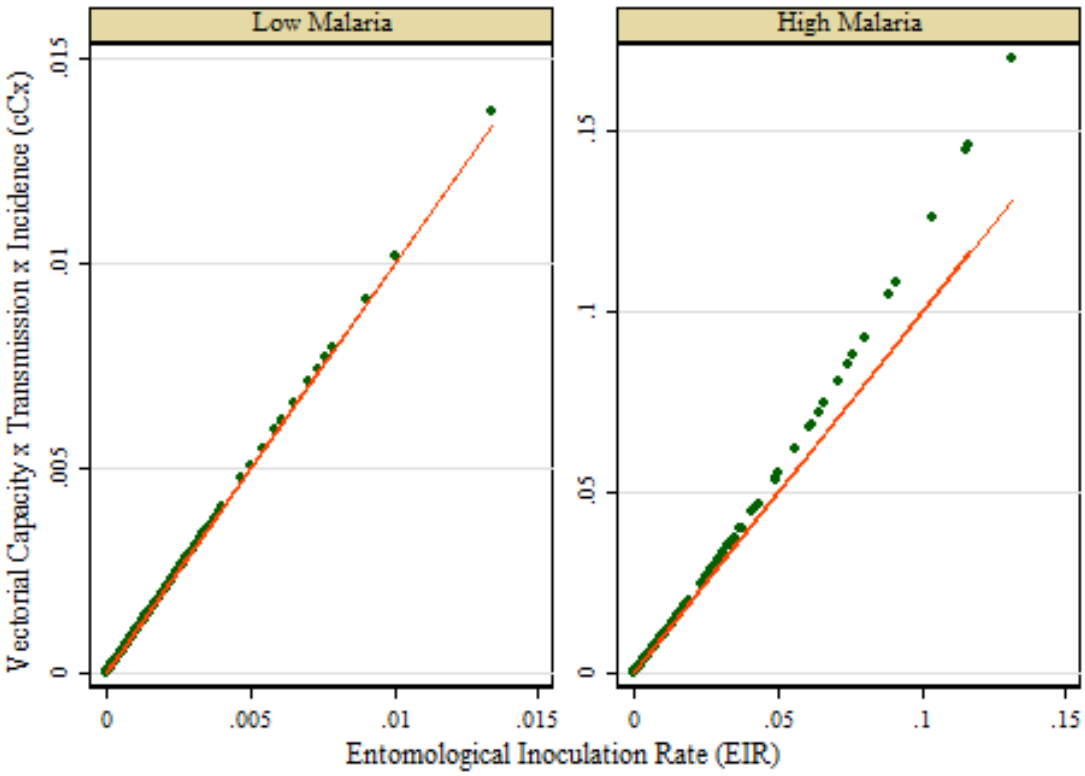


Figure 6: Testing Assumption 5 that $cCx \approx EIR$

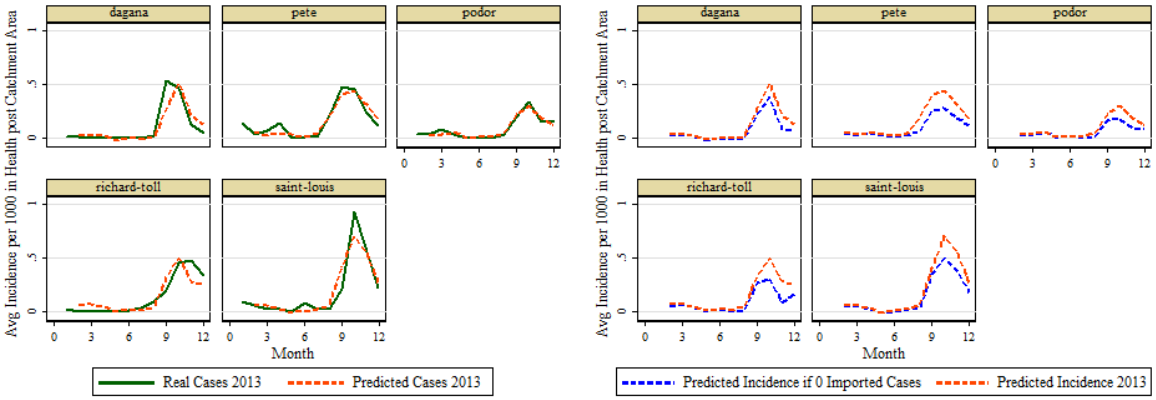


Notes: For each health post area in each month, the entomological inoculation rate is calculated as is the vectorial capacity times transmission to mosquitoes times incidence (cCx). Based on Assumption 5, in low malaria districts it should be possible to approximate EIR with cCx In the scatter plot, this means the points should line along the 45 degree line, which is shown in red. The graph shows that for the low malaria districts, this holds, but for the high malaria districts, as the incidence gets larger, the approximation is no longer accurate.

Figure 7: Predicted, Predicted without Imported Cases and Actual Incidence Averaged Across Health post Areas by District

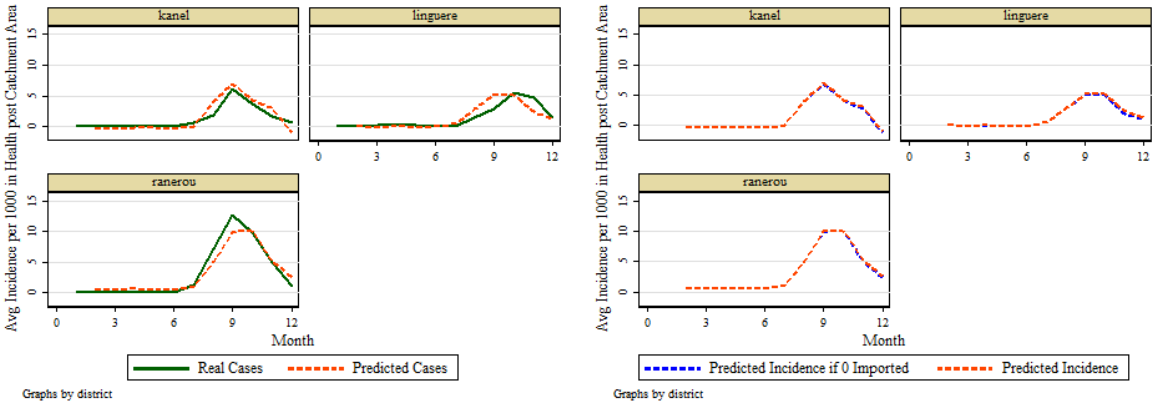
(a) Goodness of Fit, Low Districts

(b) Effect of Travelers, Low Districts



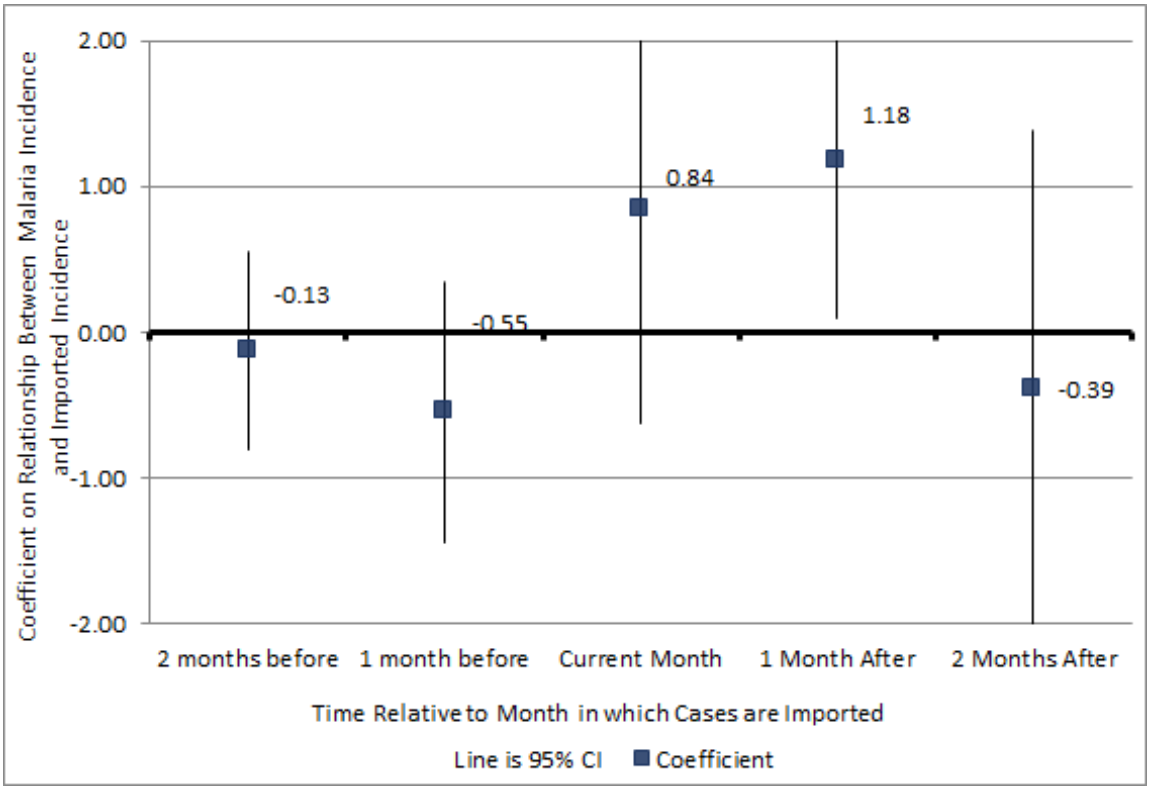
(c) Goodness of Fit, High Districts

(d) Effect of Travelers, High Districts



Notes: In all four figures, the orange dash lines represent the monthly predicted malaria incidence averaged across health post areas within a district. This was calculated based on values for the parameters of the model drawn from their distributions. I conducted 500 replications and used the mean monthly incidence value per health post area. Panels a and c compare the predicted values to the actual malaria incidence, where the solid green lines are actual incidence averaged across health posts within a district. In panels b and d, the predicted incidence is compared to the incidence that would occur if there were no cases imported by travelers, shown in dashed blue lines. Incidence with 0 imported cases was calculated using the same 500 replications for parameter values, but imported and lagged imported cases were set to 0.

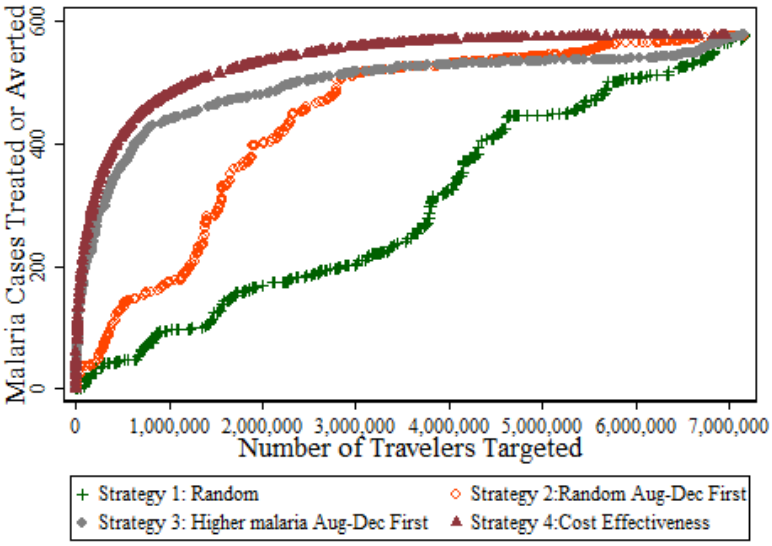
Figure 8: Estimated Impact of Future, Current and Past Expected Imported Malaria Incidence



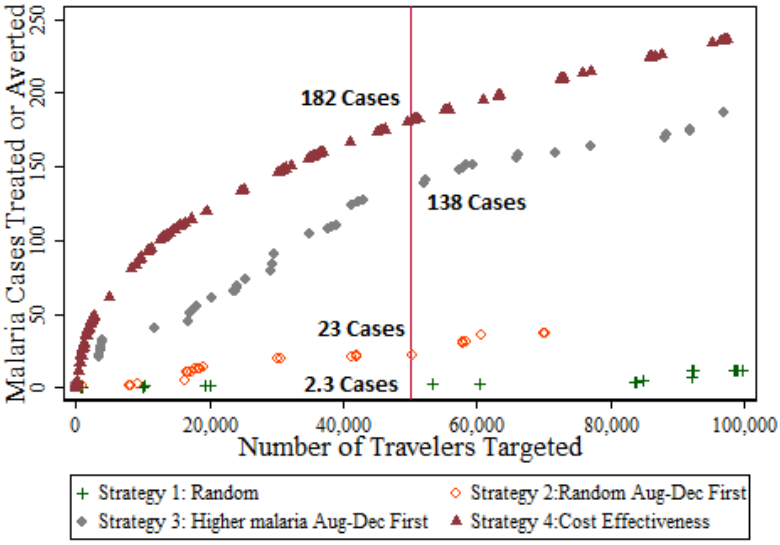
Notes: The figure was constructed based on a regression of current malaria incidence on imported incidence of malaria two months later, one month later, currently, last month and two months ago, controlling for lagged local incidence, time and location fixed effects and rainfall covariates.

Figure 9: Number of Travelers Targeted and Cases Treated/Averted Under Different Policies

(a) Full Cost Curve



(b) Cost Curve Zoomed in on Less than 100,000 Travelers Tested



*Notes:*The panels show four different strategies for targeting travelers. Each symbol represents a district-month. Targeting a specific district-month means targeting all travelers entering the five low malaria districts from that district in that month. The strategies lay out which district-months are targeted first. The cost of the strategy is the number of travelers targeted and the benefit is calculated as the number of primary and secondary cases treated or averted based on the prediction from the model estimated earlier.

Table 1: Statistics from the Cell Phone Data in 2013

Avg # of Calls/Texts per account	1657 calls/texts	
Avg # of days per account with at least one text/call	155 days	
	All Travel	Without Travel Between Neighboring Districts
% of Accounts Taking at Least 1 Trip	82%	60%
# of Accounts Traveling on Avg per Day (% of pop)	263,949 (3%)	139,940 (1.5%)
Avg Trips per Account	10 trips	5 trips
Avg # of Health Districts Visited	5 Districts	3 Districts

Notes: Travel is defined as a person changing district locations from one day to the next. Each time a person changes location, it is defined as a trip.

Table 2: Effect of Imported Malaria Incidence

Panel A: Regression Results

	(1) Low Districts Weighted and Not Scaled by Utilization	(2) High Districts Weighted and Not Scaled by Utilization	(3) Low Districts Weighted and Scaled by Utilization	(4) Low Districts Unweighted and Not Scaled by Utilization
β_3, Imported Incidence	1.161**	0.418	1.224*	1.659**
	(0.544)	(3.755)	(0.679)	(0.777)
β_2, Lag Imported Incidence	0.437*	1.086	0.609**	0.625*
	(0.246)	(2.625)	(0.295)	(0.352)
β_1, Lag Local Incidence	0.347***	0.576***	0.348***	0.347***
	(0.0581)	(0.0629)	(0.0574)	(0.0581)
Month Dummies	Yes	Yes	Yes	Yes
Health Post Area FE	Yes	Yes	Yes	Yes
Rainfall Controls	Yes	Yes	Yes	Yes
Health Post x Month Obs	396	297	396	396
R-squared	0.663	0.663	0.769	0.663
Mean Incidence	0.13	2.09	0.13	0.13
Mean Imported Incidence	0.03	0.067	0.157	0.021
Cumby-Huizinga test p-value = 0.393				

Panel B: Prediction Tests

	Low Districts
p-value of test β_3=1 (β_3 =coefficient on Imported Incidence)	0.78
p-value of test β_2=0.273 (β_2 =coefficient on Lag Imported)	0.60
p-value of test β_1=0.273 (β_1 = coefficient on Lag Local Incidence)	0.42
p-value of test β_2=β_1	0.25

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

Table 3: Primary and Secondary Cases in Richard Toll District

		(1) Primary Cases	(2) Secondary Cases	(3) # of Secondary Cases per Primary Case
2013				
	Imported	121	18	0.15
	Local	87	20	0.23
	Total Investigated	208	38	
2014				
	Imported	85	8	0.09
	Local	33	6	0.18
		118	14	
2015				
	Imported	178	20	0.11
	Local	68	8	0.12
		246	28	

Notes: Travel history data were collected by MACEPA on all positive cases of malaria and individuals within the household or geographically near households were tested to pick up secondary cases.

Table 4: Placebo Tests

	(1) Base Model	(2) Travel Scaled by Avg Monthly Incid	(3) Avg Travel Scaled by Monthly Incid	(4) Travel Scaled by All Disease Incid	(5) Effect on Non-Malarial Disease
Imported Incidence	1.213**	0.0177	0.00307	-0.00143	16.64
Imported Beta Coefficient	0.240 **	0.0156	0.00403	0.00890	.0481
	(0.492)	(0.0910)	(0.105)	(0.00772)	(21.30)
Lagged Incidence	0.350***	0.365***	0.365***	0.365***	0.245**
	(0.0561)	(0.0678)	(0.0678)	(0.0675)	(0.101)
Month Dummies	Yes	Yes	Yes	Yes	Yes
Health Post Area FE	Yes	Yes	Yes	Yes	Yes
Rainfall Controls	Yes	Yes	Yes	Yes	Yes
Health Post x Month Obs	396	396	396	396	396
R-squared	0.663	0.648	0.647	0.647	0.764

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

Table 5: Robustness Checks

Model	(1) Baseline	(2) Cases	(3) Spatial and Panel Standard Errors	(4) Non-time Varying Pop	(5) Net Imported	(6) No Neighbors
Imported	1.161**	1.295***	1.161***	1.185**	1.239**	1.725*
	(0.544)	(0.291)	(0.336)	(0.501)	(0.534)	(0.853)
Lagged Imported	0.437*	0.438***	0.437	0.424*	0.617**	0.821*
	(0.246)	(0.118)	(0.393)	(0.243)	(0.259)	(0.414)
Lagged Local	0.347***	0.347***	0.347***	0.341***	0.345***	0.376***
	(0.0581)	(0.103)	(0.0585)	(0.0600)	(0.0569)	(0.0453)
Month Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Health Post Area FE	Yes	Yes	Yes	Yes	Yes	Yes
Rainfall Controls	Yes	Yes	Yes	Yes	Yes	Yes
Health Post x Month Obs	396	396	396	396	396	396
R-squared	0.663	0.764	0.663	0.661	0.666	0.644

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

Appendices

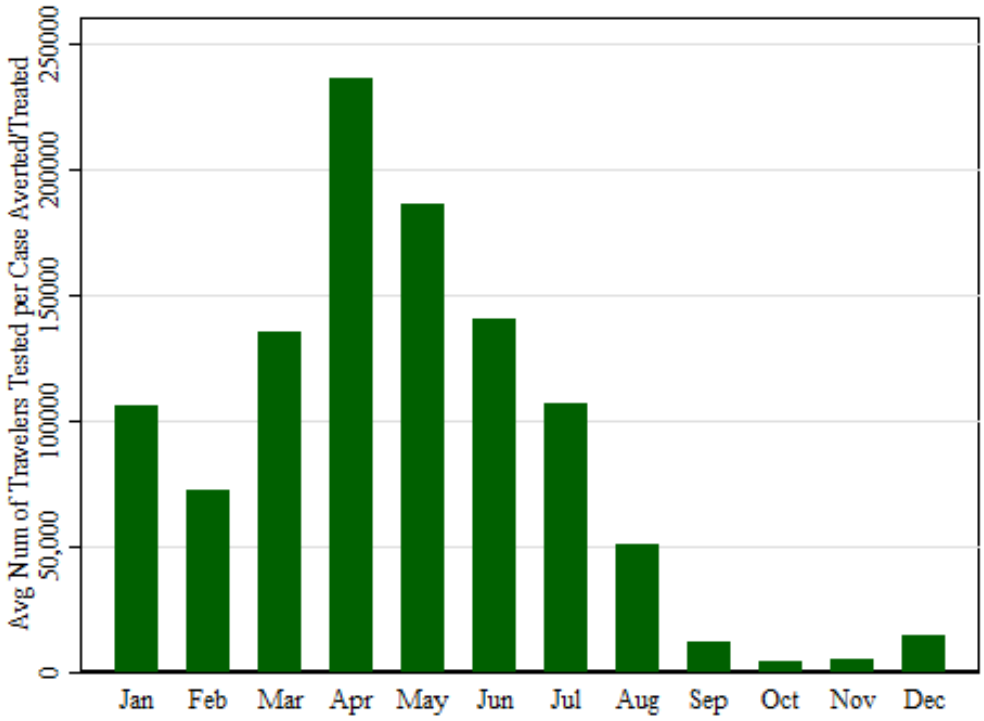
A Additional Tables and Figures

Table A1: Effect of Imported Malaria Incidence, Scaled by Proportion of Population Susceptible

	(1) Incidence Low Districts	(2) Incidence High Districts
β_3, Imported Incidence	1.161**	0.429
	(0.544)	(3.767)
β_2, Lag Imported Incidence	0.438*	1.157
	(0.246)	(2.622)
β_1, Lag Local Incidence	0.348***	0.591***
	(0.0582)	(0.0644)
Month Dummies	Yes	Yes
Health Post Area FE	Yes	Yes
Rainfall Controls	Yes	Yes
Health Post x Month Obs	396	297
R-squared	0.663	0.770

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

Figure A.1: Avg Cost Per Case in Each Month



B Primary and Secondary Case Detection in Richard Toll

In the health district of Richard Toll, MACEPA has been implementing what is known as reactive case investigation since 2012. When a patient tests positive for malaria, an investigation begins whereby within seven days detailed information, including travel history, is collected on the individual as well as every person living in the same household, and all household members are tested for malaria as well. In addition, malaria treatment is given to all household members. Finally, all individuals living in households within 100m of the initial positive case are tested for malaria, and those in a household where a positive case is found are treated for the disease. Data on all the primary and secondary cases between 2013 and 2015 are available, and provide some measure of the spread of malaria from both imported and local cases. Primary cases were only investigated when the infected individual stayed in a household within Richard Toll (and was not sleeping at the market or bus terminal, which happens with transient travelers).

The lower values for secondary transmission seen in Table 3 can be explained by the fact that the reactive case investigation provides medication to everyone in the household, regardless of the outcome of the test. Therefore, someone that might have recently been infected and does not yet have enough of the parasite to test positive using an RDT, would not be counted as a secondary case, even if he or she might have been one without the preemptive treatment.