Policy Research Working Paper 9198

Using Mobile Phone Data to Reduce Spread of Disease

Sveta Milusheva

WORLD BANK GROUP
Development Economics
Development Impact Evaluation Group
March 2020

Abstract

While human mobility has important benefits for economic growth, it can generate negative externalities. This paper studies the effect of mobility on the spread of disease in a low-incidence setting when people do not internalize their risks to others. Using malaria as a case study and 15 billion mobile phone records across nine million SIM cards, this paper causally quantifies the relationship between travel and the spread of disease. The estimates indicate that an infected traveler contributes to 1.7 additional cases reported in the health facility at the traveler's destination. This paper develops a simulation-based policy tool that uses mobile phone data to inform strategic targeting of travelers based on their origins and destinations. The simulations suggest that targeting informed by mobile phone data could reduce the caseload by 50 percent more than current strategies that rely only on previous incidence.

This paper is a product of the Development Impact Evaluation Group, Development Economics. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://www.worldbank.org/prwp. The author may be contacted at smilusheva@worldbank.org.

Using Mobile Phone Data to Reduce Spread of Disease*

Sveta Milusheva

JEL Classification: H51, I15, J18, I18, O15

Keywords: Health, big data, epidemics, mobility, public policy, mobile phones

*I am grateful to Andrew Foster, Jesse Shapiro, and Daniel Bjorkegren for their continuous guidance on this project. I would like to thank Victor Alegana, Girija Borker, David Glick, Morgan Hardy, Arianna Legovini, Emily Oster, Nick Ruktanonchai, Bryce Steinberg, Andrew Tatem and William Violette for valuable feedback. I appreciate the comments from participants at the World Bank ABCDE Conference, the NEUDC Conference, the PopPov Conference, the Population Health Sciences Workshop, the PAA Conference and the Brown microeconomics lunch seminar. Thank you to the Bill & Melinda Gates Foundation (grant OPP1114791) and the NICHHD (grant T32 HD 7338-29) for financial support. Anonymous mobile phone data have been made available by Orange and Sonatel within the framework of the D4D-Senegal challenge. I thank Philippe Guinot and Yakou Dieye, PATH and the National Malaria Control Program for providing the data on malaria incidence and helpful feedback.

1 Introduction

Increasing domestic and international mobility has magnified the devastating consequences of infectious diseases: more than 11,000 deaths from Ebola, 440,000-1,300,000 cases of Zika infections in 13 countries, and most recently, more than 125,048 infections and 4,613 deaths from COVID-19[1] across 117 countries (Bogoch et al. 2016, WHO 2020a, WHO 2020b). Negative externalities from mobility are also relevant for long-standing diseases that we are aiming to eliminate. For example, Venezuela, the first country certified by the World Health Organization (WHO) for eliminating malaria in its most populated areas in 1961, experienced a dramatic resurgence of the disease in 2016 in part due to migrant workers in the mining region becoming sick, traveling home, and spreading the disease to their home villages and cities (Casey 2016). This paper uses a case study of malaria in Senegal to demonstrate how to harness big data to causally estimate the size of this externality of movement and apply the results towards more effective policy targeting. The methods can be applied more broadly to inform policies related to mitigating spread of infectious diseases.

The economics literature identifies malaria eradication as having important impacts on adult income and consumption (Bleakley 2010, Cutler et al. 2010, Venkataramani 2012), real estate wealth (Hong 2011), longer term health including chronic disease and disability (Hong 2013), test scores and educational attainment (Barofsky, Anekwe, and Chase 2015, Lucas 2010, Venkataramani 2012).[2] Ninety-nine countries have been certified by the WHO as malaria free; however, Sub-Saharan Africa, which accounted for 93% of all malaria deaths in 2018, has only had a single successful case of elimination (*World Malaria Report* 2019). While previous work has studied malaria prevention/treatment in short-term settings, focusing on the pricing of malaria interventions (Jessica Cohen and Dupas 2010, Jessica Cohen, Dupas, and Schaner 2015, Dupas 2014, Laxminarayan et al. 2010, Tarozzi et al. 2014) as well as the adoption of preventative or anti-malarial treatment (Adhvaryu 2014, Apouey, Picone, and

[1] Data as of March 12, 2020

[2] See Currie and Vogl (2013) and Apouey, Picone, and Wilde (2018) for summaries of the literature.

Wilde 2018, Armand et al. 2017), research has not yet examined behavioral factors that may contribute to the persistence and spread of malaria in the long-term.

This paper uses novel data to causally estimate a constraining factor for elimination: the reintroduction of malaria into elimination zones by population movement. While this phenomenon has been documented in at least 61 countries by the epidemiology literature, this has not been done in a causal framework (Justin Cohen et al. 2012, Lu et al. 2014). In this research, I quantify the negative externality of mobility empirically using a similar setting in Senegal. This paper leverages policy simulations based on these estimates to show how aggregated big data on individuals' geolocation can inform more cost-effective targeting strategies to reduce transmission generated by population mobility, which would be a complementary component of a campaign to successfully eliminate malaria.

The negative externality from travel is generated when people are unaware of their risks to others because they do not know that they are disease vectors.[3] Yet given the benefits of travel, an information campaign is unlikely to cause people to internalize the externality and choose not to travel to prevent infecting others. Measuring the size of the externality and identifying those people that contribute the most can allow for targeted policies that can help address this market failure.

The main challenge in estimating the size of the externality from mobility is that while disease transmission may respond quickly to changes in migration patterns, existing survey data that record these patterns are often infrequent or do not have coverage across a country.[4] Therefore, the only strategy available to policymakers to address this externality is using incidence in the previous year to identify where and who to target. This paper is able to significantly improve on this strategy by utilizing a new source of data to track population

[3]As will be described in more detail in the next section, the long incubation period for malaria allows people to travel without knowing they are infected. Additionally, those in high malaria settings typically develop immunity and do not experience malaria symptoms yet can infect mosquitoes when they travel to low malaria settings.

[4]In Senegal, the main official source of data on population movement is census data that only includes long-term migration statistics every 10 years. There are other surveys that ask about commuting, such as the Household Mobility Survey of Dakar (EMTASUD), but it is only for one point in time, it is focused only on Dakar, and it was done in 2000 and in 2015).

movement for a large number of people between health facility catchment areas at the daily level. I leverage mobile phone metadata for 9.5 million SIM cards in Senegal in 2013 to extract patterns of movement between different areas from the approximate locations of 15 billion calls and texts. For each month and health facility area, I measure the number of incoming travelers from other regions weighted by the incidence of malaria in these regions and the length of time spent in the origin and destination to calculate "expected imported malaria cases." I study an area of Senegal close to elimination to focus on reintroduction effects. I use a panel data strategy to estimate the impact of imported incidence on total malaria incidence in this low-malaria setting using a linear dynamic panel-data model and controlling for time fixed effects. If infected travelers only lead to a malaria case being detected in the destination rather than the origin of the traveler, but do not generate any externality in the form of additional malaria cases, then a standard model would predict for each expected imported case one more additional case reported in the destination. Instead, I find that one additional expected imported case of malaria in a low malaria area leads to 1.7 cases of malaria reported, indicating an externality of .7 new cases.

Given that migration has numerous economic and social benefits, policymakers face trade-offs between economic growth and improving public health in designing policies to reduce travel-linked malaria cases. This paper provides a useful framework for strategic targeting of high-risk populations in low-incidence areas to reduce negative externalities from travel with minimal interference to travel patterns. There are two categories of targeting considered: (1) targeting high-risk travelers entering a low malaria area from a high malaria area and (2) targeting all travelers in only specific areas of low-malaria regions that are likely to receive many high-risk travelers.[5] Within each type, I compare a strategy that incorporates daily information on origins and destinations of travelers from mobile phone data with strategies that only use information on incidence in the previous year that could be implemented by the government in the absence of mobile phone data. The most cost-effective

[5]This paper does not focus on the type of targeting, but examples can include information campaigns targeted to travelers via mobile phone and strategically setting up testing sites.

strategy is to use mobile phone data and combine the two types of targeting. On average, given the existing budget available for this type of activity, the cost- effective strategy using mobile phone data performs over 50% better compared to the next best strategy that only relies on incidence in the previous year.

My empirical design accounts for confounders correlated with movement. Using rainfall proxies and month fixed effects, I control for seasons and holidays, which drive a large amount of migration in Senegal. I also test that it is not an unobservable correlated with both migration and malaria driving the results, but instead the combination of movement and the malaria levels at origins and destinations. Additional checks show that the relationship between imported incidence and malaria incidence is not driven by some other relationship between origins and destinations as well as to ensure that the relationship holds only for malaria and not for other health conditions. I also test the impact of future imported incidence on malaria incidence in the current month and find no relationship.

This paper builds on previous health literature that has established travel as a risk factor for contracting malaria, (Montalvo and Reynal-Querol 2007, Lynch et al. 2015, Osorio, Todd, and Bradley 2004, Siri et al. 2010, Littrell et al. 2013), by estimating the size of the causal impact from an expected imported case, which makes it possible to conduct policy simulations and compare different targeting strategies. Tatem et al. (2009) and Le Menach et al. (2011) use three months of cell phone data to estimate the malaria importation rate to Zanzibar using a static model that does not account for seasonality due to the limited time frame of their mobile phone data. Similarly, Wesolowski, Eagle, et al. (2012), Enns and Amuasi (2013), Chang et al. (2018) and Ihantamalala et al. (2018) among others, do not incorporate seasonality in incidence and focus on identifying potential sources and sinks based on travel patterns and annual malaria prevalence data. Yet Buckee, Tatem, and J. Metcalf (2017) point out that seasonal variation in biological factors related to climate and seasonal population movements are important for many infectious diseases and failing to account for seasonality could lead to misallocation of resources.

While Wesolowski, Erbach-Schoenberg, et al. (2017) look at seasonality of movement patterns across Kenya, Pakistan and Namibia, they only connect this theoretically to impact on disease and do not study the relationship with incidence data. Papers that have combined seasonal mobility data from mobile phones with seasonal disease incidence data, such as Wesolowski, C. Metcalf, et al. (2015) for rubella and Wesolowski, Qureshi, et al. (2015) for dengue, have not done so in a causal framework. This paper contributes to the existing work by aiming to measure the causal relationship and size of the effect of imported malaria using a linear dynamic panel-data model and controlling for time fixed effects. Therefore, in addition to the two areas already identified by the economics literature as necessary for malaria reduction—pricing and adoption of preventative and treatment interventions—this paper identifies targeting of higher risk mobile populations as a third.

While this paper focuses on malaria elimination, it has implications for other diseases whose spread has been associated with travel (Adda 2016, Oster 2012,Prothero 1977, Balcan et al. 2009, Stuckler et al. 2011, Tam, Khan, and Legido-Quigley 2016). Since travel patterns studied using 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 methods developed in this paper. I demonstrate how new sources of big data can be used to measure externalities associated with travel to develop more effective targeting strategies that can be combined with pricing and adoption policies. This further expands the use of big data for development in areas such as risk-sharing (Blumenstock, Eagle, and Fafchamps 2016), measuring poverty (Blumenstock, Cadamuro, and On 2015, Blumenstock 2016) and providing credit to the poor (Bjorkegren and Grissen 2018).

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 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. The malarial cycle for *P. falciparum*, the parasite causing 100 percent of cases in Senegal, can take several weeks (*World Malaria Report* 2014). 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).[6] 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.[7] Combining the two incubation periods, a secondary case will take around one month to appear after a primary case.[8]

This paper focuses on the role of human behavior on spread of the disease.[9] There are two channels through which population movement can lead to spread of malaria in low-malaria or elimination zones. The first is residents of these zones 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.Therefore if they are bit by a mosquito in the low malaria area, they could infect that mosquito and it could in turn infect other individuals.

[6]The incubation period can vary, but two different sites in Senegal had an average of 9 days.

[7]Unlike other malarial parasites, *P. falciparum* does not have the potential to lie dormant for months.

[8]Details on malaria transmission can be found in Doolan, Dobaño, and Baird 2009, D. L. Smith and McKenzie 2004, Killeen, A. Ross, and T. Smith 2006, Wiser 2010.

[9]Average radius of travel for the mosquitoes that carry the malaria parasite in Senegal is only 1-2 km; therefore, mosquito movement is not considered (Russell and Santiago 1934, Thomas, Cross, and Bøgh 2013).

2.2 Health System and Malaria in Senegal

Senegal is geographically divided into 14 health regions, under which there are 76 health districts. The main point of service for malaria cases is the health post. There are a total of 1,247 health posts in the country (PNLP, INFORM, LSHTM 2015). In addition, there are rural health huts and community health workers that provide care for those living far from a health post, and report the cases to the closest health post.

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 (PNLP, INFORM, LSHTM 2015). Currently, the north of the country has very low incidence and is at the level considered ready for elimination by the World Health Organization (1 case per 1000, 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.[10] 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. Nevertheless, the mosquitoes required to spread the disease are also present in the low malaria areas (Ndiath et al. 2012). Given the two distinct zones in the country, the Government of Senegal strives to continue reducing the case load in the South, while aiming to eliminate it completely from the North. As potentially infected individuals travel from the South to the North, though, they can hinder elimination efforts in the North.

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).[11] A large part of this migration is rural to urban due to irregularity of rainfall and degradation

[10] The Appendix contains a map of annual malaria incidence by district.

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

of the ecosystems that have impacted agricultural activity (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 as people return home to visit family and friends or receive visitors from home (Cho, Myers, and Leskovec 2011). Focusing on migrants in Dakar, A. S. Fall (1998) finds that 87% of male and 81% of female migrants visited their home areas, with the majority of visits occurring for holidays, family ceremonies and religious festivals.

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). Broader 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 (Herrera and Sahn 2013). Additionally there are still pastoral groups that travel within a set territory(Adriansen 2008). 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. These factors and their relationship to malaria incidence will be discussed in the model section.

In Senegal, 2% of the population are international migrants while only 1.2% of the population emigrated from Senegal. Focusing on immigration into Senegal in 2013, only 0.23% of the population entered the country. While the paper focuses on the role of internal migration, the potential impact of international migration will be discussed.

2.4 Malaria Data

Low-incidence areas close to elimination can experience the largest externality from population mobility for three key reasons: (1) without these travelers the disease could be reduced to zero and require lower government expenditures; (2) in high malaria areas, people have usually built up an immunity to the disease; therefore, a traveler entering a high malaria area is less likely to lead to a new infection even if he or she infects additional mosquitoes

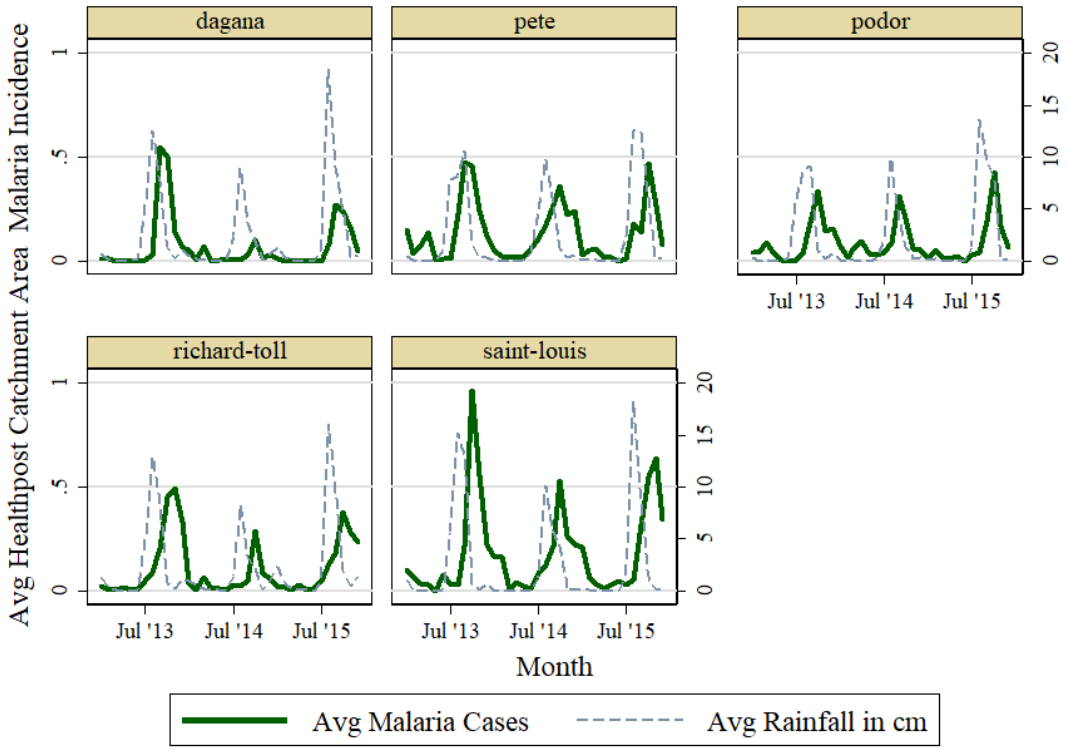
in the area, while in a low malaria area immunity does not exist; and (3) the infection in a low-malaria area is likely to be more severe due to the lack of exposure to the disease. Therefore, I focus on the part of Senegal discussed earlier that is at a pre-elimination stage. Within this area, I focus on five of the lowest malaria districts where data are disaggregated at the health post level and available for every health post in these districts. Malaria data are not available at this high spatial resolution for any of the other low malaria districts. The data cover 117 health posts. The appendix provides a map of the five health districts, which I subdivide into areas based on the location of the health posts and cell phone towers. Health posts in close proximity were grouped together forming 36 health post catchment areas.

I use incidence data based on data collected from each health post on all new cases in the reporting month.[12] The use of incidence data is one thing that separates this paper from some of the previous work that relies on endemicity data. 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, the likelihood of having a positive case becomes very small. Therefore, when focusing on a low-malaria setting to understand impact of mobility, incidence is a more reliable measure (Alegana et al. 2013, J. M. Cohen et al. 2013).

PNLP's work has led to a system that provides high quality data on malaria incidence across the country. In Senegal, if an individual feels sick, usually experiencing a fever, chills and fatigue, she will go to the closest health post where she will be tested using a rapid diagnostic test (RDT) due to her symptoms. If she tests positive, she will be provided with medication for free to treat the disease. Therefore, all incidence data used in this paper comes from suspected cases that have been tested and are positive for malaria based on the test. For the rest of the country, these incidence data are available monthly at the health

[12]The data used to measure malaria incidence comes from the PNLP and PATH, a non-profit organization working with the PNLP to fight malaria in Senegal through its Malaria Control and Elimination Partnership in Africa (MACEPA).

Figure 1: Average Monthly Health post Catchment Area Malaria Incidence per 1000 and Average Monthly Rainfall, Jan 2013-Dec 2015 by Health District



district level. These data are used to classify the risk of travelers based on their origin.

Monthly malaria incidence per 1000 people is averaged across health post catchment areas within districts for three years in Figure 1. Districts on average have around 0.1 cases per 1000 people per month. The figure overlays the monthly cumulative rainfall in centimeters averaged across health post catchment areas.[13] The comparison of cases and rainfall demonstrates strong seasonality of malaria in Senegal and the close relationship between rainfall and malaria, with the peak of cases annually occurring one to two months after the peak in rainfall. I model this relationship in the analysis since rainfall can be correlated with both malaria and population movement.

There are three main challenges that arise with using clinical data: incomplete data reporting, presumptive diagnosis based on symptoms rather than testing and non-utilization

[13]Rainfall data are from the Climate Prediction Center (2016) Rainfall Estimator for Africa.

of the public health system (Alegana et al. 2013). In the data only 12 out of 1,416 health post-month observations are missing in 2013. In addition, 99% of suspected cases were tested parasitologically in the five districts analyzed. Since both malaria cases and imported cases are calculated based on case data, as long as utilization is relatively uniform across the country, it should not bias results. 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. While in the main analysis, I assume uniform utilization, I include a robustness check where cases are scaled by regional utilization in the DHS.

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 (Montjoye et al. 2014). The data come from the second phase of the Challenge and consist of 15 billion call and text records for Senegal between January 1, 2013 and December 31, 2013 for all of Sonatel's user base.[14] The data contain information on all calls and texts made or received by a SIM card, their time, date and location of the closest cell phone tower, which enables tracking of SIMs in space as they make calls from different tower locations. The data are anonymized, with a random ID provided that makes it possible to track the same SIM over time, but no identifying information on the individuals. On average there are 1,657 calls or texts per ID during the year, and on average an ID has a call or text on 155 days.

Each tower is assigned a health district based on its GPS coordinates. I follow previous literature to assign individuals a daily health district location based on the cell tower of the last call or text of the day (Ruktanonchai et al. 2016). In instances where there are days with no calls, I replicate Wesolowski, Eagle, et al. (2012) and assign the health district location of the day closest to the one missing.[15] A health district location is assigned to each SIM

[14] At this time it was not possible to obtain more recent data or data from other providers.

[15] The appendix includes a robustness check where observations with more than 14 days in a row missing

for every day of the year.

Movement is defined as a change in location from one health district to another between two consecutive days. The population is highly mobile, with over 80% of all Sonatel SIM cards taking at least one trip and over a quarter million traveling on average on any given day. On average annually per SIM there are 10 different trips to almost five different health districts. In addition to assigning the towers within the study area to a health district, I assign them to a health post catchment area based on their GPS location. Therefore, each traveler entering one of the five health districts is assigned a specific health post catchment area based on the last call or text of the day.[16]

Panel a of Figure 2 shows the average number of people entering a health post catchment area each day as a percent of the population in that catchment area averaged across all areas, along with vertical lines marking several religious holidays and important pilgrimages. The movement patterns largely align with the holidays and pilgrimages, which supports 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 of that area enters on any given day. The variation in percent of people entering can vary widely by health post catchment area and date within a district (Panel B). 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.

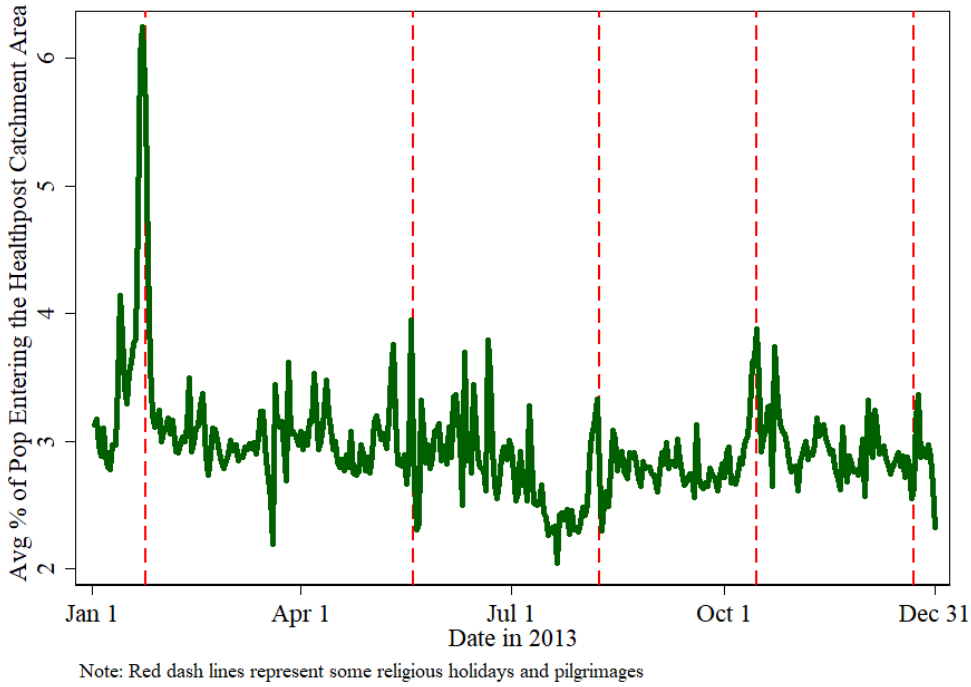
In 2013, Sonatel had slightly over 9.5 million unique phone numbers on its network while the population of Senegal was 13.5 million.There are two sets of people that are potentially excluded from the data and need to be accounted for–those without a phone and those with

are removed.

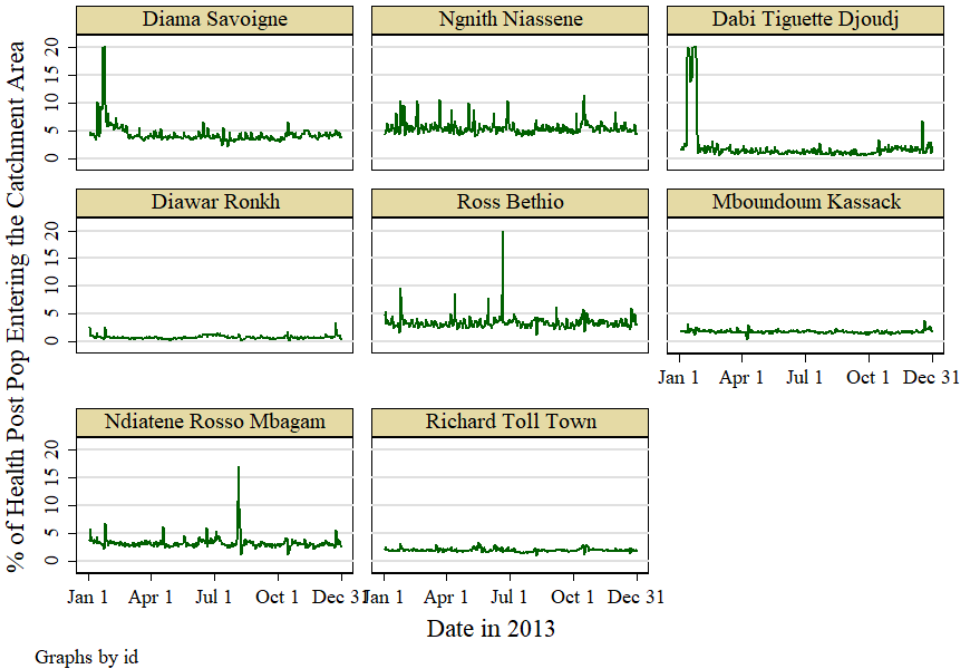
[16]Movements within a district between health post catchment areas are not counted.

Figure 2: People Entering a Health Post Catchment Area as a Percent of the Population in the Area

(a) Avg Across All Health Post Catchment Areas



(b) Health post Catchment Areas in Richard Toll



Notes: Red dash lines in panel (a) represent some important religious holidays and pilgrimages. The scale in panel (b) is bounded at 20%, but for Diama Savoigne and Dabi Tiguette Djoudj, the value in January goes up to around 50%.

a phone but using a different mobile provider. Based on the Listening to Senegal Survey (LSS) done in 2014, 12.3% of adults age 18 and over never use a mobile phone. Sonatel is one of three mobile phone providers. Based on the LSS, Sonatel is the main provider for 80% of those surveyed with a cell phone, and 88% of those with a cell phone have a Sonatel SIM card (Agence Nationale de la Statistique et de la Démographie - Ministére de l'Economie 2014). Therefore, only around 12% of adults with a SIM are excluded. Combining the two types of missing adults, 77.2% of adults are represented in these data.

I conduct checks to see how representative the data are for the two types of users that are not included—those without a SIM and those with a SIM from a different provider. I use the DHS survey to compare mobility patterns between women with and without a cell phone in the household.[17] 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. Using the LSS, I compare several indicators between people that have a Sonatel SIM card and those that only have a SIM card from another provider. Based on t-tests, there is no significant difference between these categories of individuals based on type of primary activity (p-value=0.344), sector of primary activity (p-value=0.863), source of drinking water (p-value=0.868), non-food expenditures over past month(p-value=0.193), and non-food expenditures over past 12 months(p-value=0.115). The third missing group is children. If children travel, they are likely traveling with adults (though they might be traveling less on average if they do not always travel every time an adult travels). 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 represent the full population and to get an upper bound on movement. The weighted data overestimates total movement and underestimates the impact of each trip. Unweighted results are provided in the robustness section as an upper bound on the effect size.

[17] The survey does not contain data on men.

3 Empirical Model

The empirical specification is derived from a model of malaria that is based on previous models used in D. L. Smith and McKenzie (2004), Cosner et al. (2009) and Torres-Sorando and Rodriguez (1997). Four key assumptions allow me to simplify the model so that incidence in the current month is dependent on incidence in the last month using a linear functional form. Expected imported cases enter as a linear additive term as in Torres-Sorando and Rodriguez (1997). The model, assumptions and implications are described in detail in Appendix A. I start out estimating equation 1 using OLS, with imported incidence calculated using equation 2:

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

$$\mathbb{E}(\mathcal{I}_{it}) = \frac{1}{H_{it}} \sum_{j \neq i} \sum_{p_t \in j} T_{ip}(x_{jt} T_{jp}) \quad (2)$$

In this model, x_{it} represents the incidence, or number of humans infected in location i at time t per 1000 people in location i.[18] Location i is one of the 36 health post catchment areas and t is at the monthly level. The β_1 parameter estimated tells us on average how many new cases per 1000 are generated in the following month from cases in the current month. In contrast to epidemiological models, where this parameter would be estimated separately for each area, I want to causally estimate the secondary cases generated. Therefore, I estimate an average effect across locations and time in order to be able to include health post area fixed effects, γ_i and month fixed effects, δ_t. This allows me to control for malaria seasonality and unobservable characteristics of a health post area that might impact incidence.

I use the mobile phone data to calculate expected imported malaria cases entering and divide them by the population of the area they enter, H_{it}, to calculate the expected imported incidence, $\mathbb{E}(\mathcal{I}_{it})$. The likelihood of an infected case entering depends on the origin of an

[18] Population is based on the known population of each health facility catchment area. The appendix includes a robustness check where the annual population is adjusted monthly using the number of people entering and leaving each month based on the mobile phone data.

individual, how long the person spent there and the length of time in the destination. I make two assertions:

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

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

The contribution of person p_t, who enters i from j at time t, to imported cases is calculated based on the incidence of the origin district and the length of time spent in origin district j and destination health post catchment area i. Only up to 15 days in the origin and up to 15 days in the destination are considered since the human incubation period is 15 days. Therefore, T_{ip} is the proportion of 15 days p_t spends in i after entering i and T_{jp} is the proportion of the month up to 15 days that p_t spent in j with monthly malaria incidence x_{jt} in month t.[19]

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 destination and the origin, and the population in the destination. Based on a partial R^2 of .44 the month explains a large portion of the variation, while .33 of the variation is explained by the health post catchment area. 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 shows how the identification comes from the unique combination of detailed data on travelers and the incidence in the origin. The matrix Z_{it} includes zero, one and two lags of rainfall, which capture both the agricultural seasons that could influence movement and changes in malaria incidence due to

[19] 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 is infected.

environmental factors. Additional functional forms of rainfall were also tested but did not significantly change the analysis; therefore, a linear functional form was used for rainfall.[20] ϵ_{it} represents idiosyncratic shocks. I cluster errors at the health post catchment area level to account for the fact that errors are correlated within panels.[21] The main coefficients of interest are β_1 and β_2, which represent the number of secondary cases generated by infected travelers and the number of primary malaria cases imported by infected travelers.

3.1 Identification

For my identification to be correct, it is necessary that within a health post catchment area over time, any idiosyncratic shocks in malaria incidence are not correlated with expected imported malaria incidence. Agricultural seasons and holidays are the two major reasons for travel. Agricultural seasons are correlated with rainfall, and additionally, rainfall could affect the conditions for travel (quality of roads). I control for this potential confounder by including rainfall covariates in my specification.

It is also possible that holidays, which increase population movement, could affect malaria. People might spend more time outside during the holidays and be exposed to mosquitoes. I address this potential threat to identification using a placebo test where I scale travelers by average monthly incidence in the country rather than by the incidence of their origin. I also examine the relationship between past and future imported cases and current malaria.

Finally, the dynamic panel model with a relatively short panel of 12 time periods could introduce a bias if the error term is mechanically correlated with the lagged dependent variable on the right hand side (Nickell 1981). I study this by comparing the fixed effects model with a random effects model. Given the results of this comparison, the preferred specification used is an augmented version of the Arellano-Bond Generalized Method of Moments estimator designed to address situations with "small T, large N" panels.

[20] The appendix includes results with these different specifications.

[21] I include a robustness check with spatial and panel autocorrelated standard errors.

4 Results

4.1 Quantifying the Effect of Imported Cases

Each imported case of malaria is associated with 1.23 cases of malaria in the current period and 0.330 cases in the next period based on the fixed effects model (Column 1 of Table 1). This specification assumes the externality from locally generated and imported cases will be the same. I explicitly test this by including lagged imported incidence along with lagged non-imported incidence (Column 2). The coefficient on lagged imported incidence is not significantly different from the coefficient on lagged local incidence, which implies that there is no differential effect between lagged imported and lagged local incidence.

I estimate a random effects model to test if there could be a dynamic panel bias due to the inclusion of fixed effects with a relatively short panel (Column 3 of Table 1). The coefficient on imported incidence is smaller, while the coefficient on lagged incidence is larger. In using random effects, though, I am no longer controlling for time-invariant characteristics of the health post areas that could be correlated with both imported incidence and malaria incidence. I include several characteristics of the health facility areas, including population density, a dummy for urban areas, and a dummy for health facility areas that are not along the border of the country (Column 4). Including these covariates, the coefficient on imported incidence is bigger and closer to the coefficient from the fixed effects model.

A Hausman test comparing the two models finds they are significantly different. Given each model has potential to be biased, since the fixed effects model might have some dynamic panel bias while the random effects model might have omitted variable bias, I use an Arellano-Bond specification (Column 5 of Table 1). Based on this model, for each imported case of malaria per 1000, there are 1.09 cases per 1000 reported. In addition, for each lagged case per 1000, there is an additional 0.563 of a case generated the following month. This also represents the negative externality of an imported case the previous month.

The epidemiological model that the empirical specification is based on leads to several

Table 1: Effect of Imported Malaria Incidence

	(1) Fixed Effects	(2) Fixed Effects	(3) Random Effects	(4) Random Effects	(5) Arellano Bond
Imported Incidence	1.230**	1.175**	0.793**	0.864**	1.094***
	(0.452)	(0.476)	(0.398)	(0.381)	(0.357)
Lag Incidence	0.330***		0.434***	0.418***	0.563***
	(0.0509)		(0.0579)	(0.0552)	(0.129)
Lag Imported Incidence		0.449*			
		(0.231)			
Lag Non-imported Incidence		0.340***			
		(0.0571)			
Rain in cm	0.00639	0.00592	0.00348	0.00449	-0.000654
	(0.00934)	(0.0102)	(0.00904)	(0.00853)	(0.00790)
Lag Rain in cm	0.0292	0.0294	0.0294*	0.0300*	0.0269
	(0.0189)	(0.0188)	(0.0177)	(0.0177)	(0.0181)
Lag 2 Rain in cm	0.0381**	0.0372**	0.0376***	0.0382***	0.0319**
	(0.0158)	(0.0155)	(0.0138)	(0.0141)	(0.0156)
Constant	-0.0381	-0.0396	-0.0487**	-0.0621**	-0.0688**
	(0.0294)	(0.0317)	(0.0237)	(0.0264)	(0.0290)
Month FE	Yes	Yes	Yes	Yes	Yes
Health Post Area Controls	No	No	No	Yes	No
Health Post x Month Obs	432	396	432	432	432
R-squared	0.509	0.512			
Hausman Test Comparing Column 1 and Column 4				p-value=0.0083	
Testing Predictions					
Test Lag Imported =Lag Non-Imported				p-value=0.650	
Test Imported=1				p-value=0.792	
Test Lag Incidence=0.273				p-value=0.0247	
Test Imported + Lag Incidence=1				p-value=0.0428	

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

Notes: Column 4 includes controls for health post area population density, a dummy for urban health post areas and a dummy for non-border health post areas.

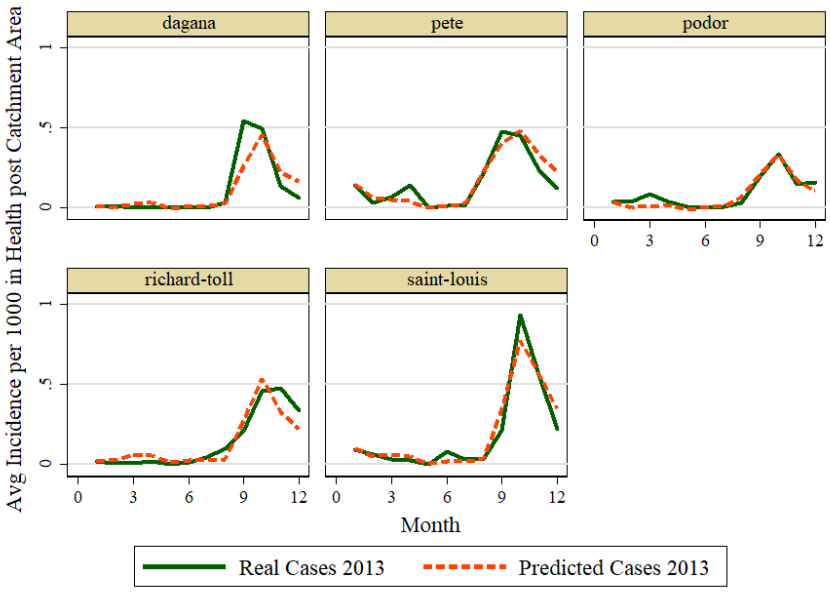
testable predictions. Results are at the bottom of Table 1. The model implies that each imported case per capita should contribute a case per capita to the incidence in the destination in the month entered. With a p-value of 0.79 on the Wald test, the coefficient is not significantly different from 1. Additionally, the β_1 parameter estimated represents the expected number of humans infected per infected human per day if transmission efficiency were perfect, C, times the actual transmission efficiencies from infected mosquitoes to humans and humans to mosquitoes, b and c. Using estimates drawn from the epidemiology literature, an average monthly value of 0.273 is expected for bcC.[22] I find that the coefficient is significantly different and larger, implying that the parameters estimated by the epidemiological model would underestimate the externality. Finally, to measure whether there is an externality beyond a case being detected in one location versus another location due to travel, I test whether the sum of imported incidence and lag incidence is equal to 1. Their sum is significantly different from 1 at the .05 level.

I simulate a baseline scenario based on the preferred model using Arellano-Bond and compare it to the actual data. I draw values for the parameters of the model from their estimated distributions, which are estimated using a bootstrap where I resample observations in the data 10,000 times. I use these draws to calculate an incidence path for the year and average paths across 500 draws. In Figure 3 panel a, I average across health post areas by district and compare to actual incidence. Similarly, I simulate incidence under the assumption that there is no imported incidence. Panel b of Figure 3 compares the results of this simulation to the baseline simulation. Imported incidence represents a substantial portion of the malaria incidence, especially for Richard Toll and Saint-Louis. On average annually per health post, travel represents 41% of the incidence.

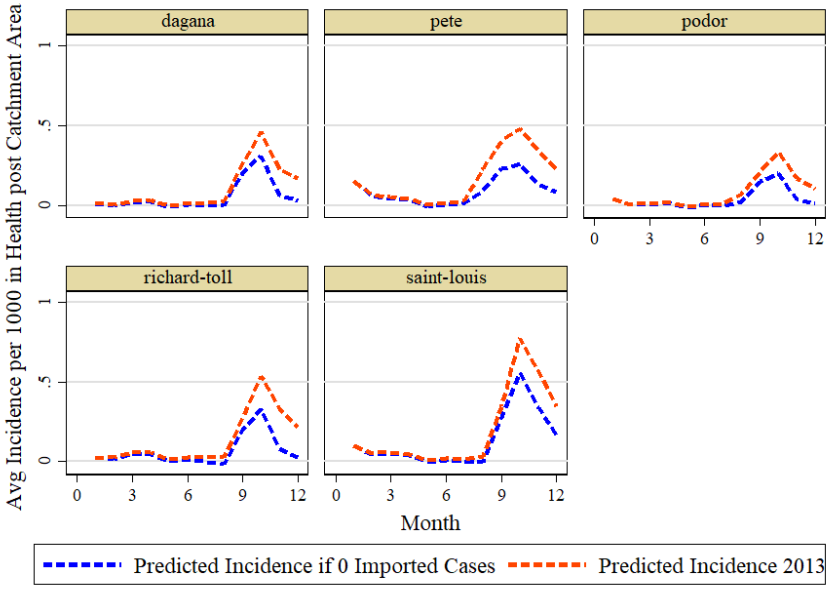
[22] Based on Vercruysse, Jancloes, and Van de Velden (1983) the average daily vectorial capacity in Senegal, C, is 1.13. Based on Gething et al. (2011), average transmission from infected humans to mosquitoes, c, is 0.161. The transmission from mosquitoes to humans, b, is 0.05 based on the linear model in D. L. Smith, Drakeley, et al. 2010. This gives an average daily value of 0.0091, or a monthly value of 0.273 for bcC.

Figure 3: Predicted, Predicted without Imported Cases and Actual Incidence Averaged Across Health post Areas by District

(a) Goodness of Fit



(b) Effect of Travelers



Notes: 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. Panel a compares the predicted values to the actual malaria incidence, where the solid green line is actual incidence averaged across health posts within a district. In panel b, the predicted incidence is compared to a scenario where no cases were 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 cases were set to 0.

4.2 Mechanism Evidence

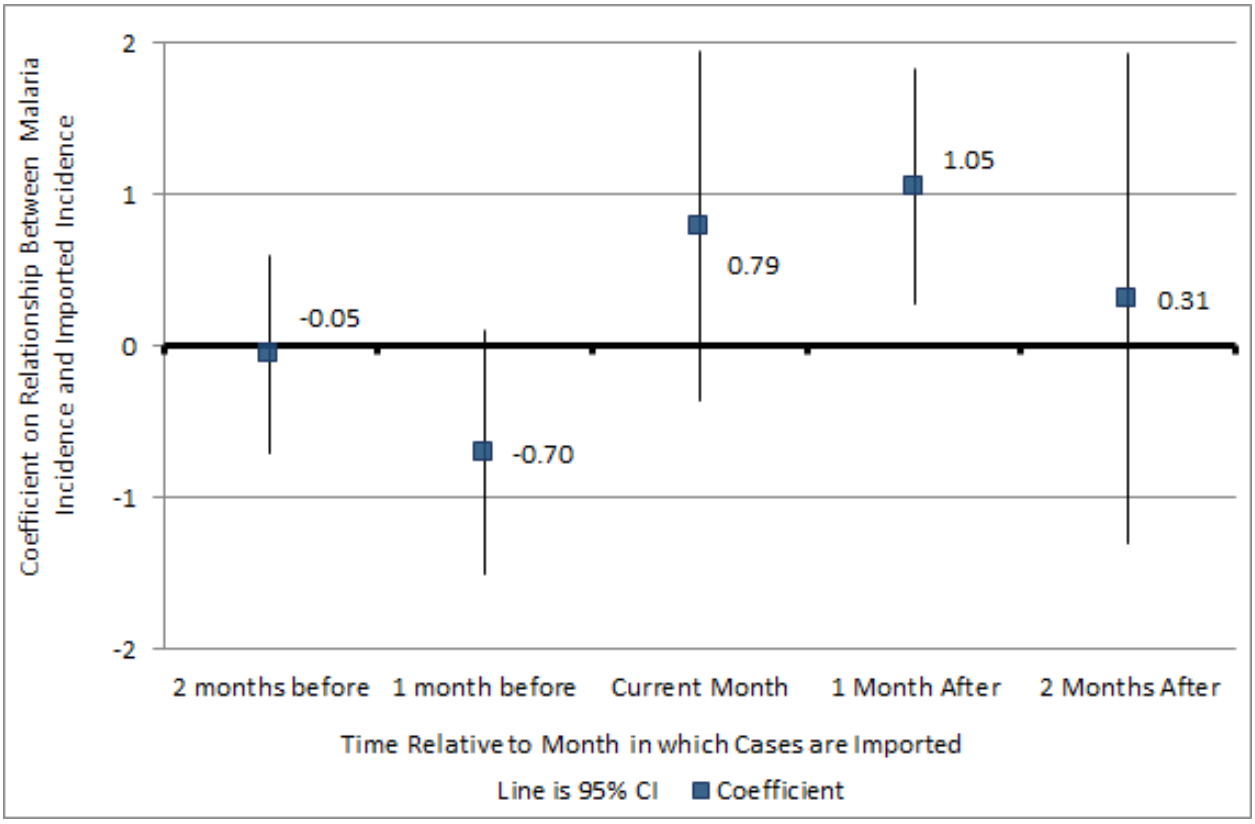
I now provide some additional evidence of the causal impact of imported incidence on total incidence. Figure 4 shows coefficients from a regression of malaria incidence on two lags and two leads of imported incidence, along with location and time fixed effects and rainfall controls. Malaria incidence 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 are associated with imported incidence, as both of those coefficients jump. Imported incidence does not seem to have a significant impact on incidence two months later.[23] Since the sample size is much smaller after including the leads and lags, the standard errors are larger. Nevertheless, the trend in the size of the coefficient still demonstrates that future imported incidence does not drive current malaria incidence.

I conduct several placebo tests to refute potential alternative explanations and compare them to the main specification (Table 2). One alternative explanation is that imported incidence is correlated with periods of higher travel such as religious holidays, and these holidays could also be positively correlated with malaria for reasons unrelated to travel. During 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 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. There is no relationship between this alternative variable and malaria incidence (Column 2, Table 2).

It is possible that only the incidence of the origin matters. For example, if the origin of travelers has a similar incidence to the destination, the variable of interest may capture this correlation between origin and destination irrespective of travel. To test this, I calculate expected imported incidence based on the incidence of the origins. Rather than separately

[23] I would not expect a persistent effect of imported cases two months out because the malaria season when the necessary mosquito vector is present in these areas is very short, lasting only around 3 months; therefore, there is not enough time for tertiary cases to develop due to the month long incubation periods.

Figure 4: 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 time and location fixed effects and rainfall covariates and clustering errors at the health post area level.

Table 2: Placebo Tests

	(1) Baseline Model	(2) Travel Scaled by Avg Monthly Incid	(3) Avg Travel Scaled by Monthly Incid	(4) Travel Scaled by Non-Malarial Case Incid	(5) Effect on Non-Malarial Case Incid
Imported Incidence	1.094***	0.0290	0.00118	0.323	27.14
	(0.357)	(0.0566)	(0.0579)	(0.337)	(25.09)
Lag Incidence	0.563***	0.588***	0.588***	0.574***	0.202*
	(0.129)	(0.130)	(0.130)	(0.144)	(0.116)
Rain in cm	-0.000654	-0.00530	-0.00538	0.0444	-0.605
	(0.00790)	(0.00741)	(0.00733)	(0.0560)	(0.843)
Lag Rain in cm	0.0269	0.0290	0.0288	0.0595	-0.198
	(0.0181)	(0.0179)	(0.0176)	(0.0564)	(0.640)
Lag 2 Rain in cm	0.0319**	0.0382**	0.0381**	0.0542	-1.340**
	(0.0156)	(0.0160)	(0.0161)	(0.0399)	(0.559)
Constant	-0.140	-0.389	-0.375	0	0
	(0.161)	(0.269)	(0.263)	(0)	(0)
Month FE	Yes	Yes	Yes	Yes	Yes
Health Post x Month Obs	432	432	432	432	432

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

Notes: Augmented version of the Arellano-Bond Generalized Method of Moments estimator used in all specifications. Columns 2-4 show regressions where imported incidence has been constructed in an alternative way that does not incorporate both incidence in the origin and the timing of the particular traveler. In Column 5, the dependent variable is non-malarial disease incidence while imported malaria incidence is calculated as is done in the baseline specification.

calculating a probability for each person using the length of travel, I use the average number of travelers, and average time spent in place of origin and destination.[24] When imported incidence is calculated for an average number of travelers per destination/origin pair, there is no longer a relationship between this variable and malaria incidence (Column 3, Table 2). These two tests demonstrate that the interaction of location, number of travelers and time spent is necessary for the significant relationship with malaria incidence.

I test that the effect is only seen for malaria incidence and is not just a reflection of a relationship between migrants and consultations at a health post. I use a variable for imported health caseload other than malaria. This variable is based on number of consultations at a health post, including all consultations with a nurse, all maternity consultations (excluding pre-natal) and all post-natal consultations. Any cases of malaria or fever are removed from the total, and the variable is scaled by the population of the health post area. This variable has no relationship with malaria incidence in the location where people enter (Column 4). I also use the original imported malaria incidence variable as the regressor, but I change the dependent variable to be non-malarial caseload as proportion of population, and I use lagged caseload instead of lagged malaria incidence. Imported malaria incidence does not have a relationship with non-malarial caseload (Column 5).

5 Policy Targeting Strategies

I study effective allocation of resources to mitigate the negative externality of travel quantified in the previous section. I first describe some of the existing policies implemented in Senegal towards travelers. Then, I conduct simulations to demonstrate the effectiveness of strategically targeted policies.

[24] Average number of travelers was calculated based on total number of travelers and number of unique origin destination pairs in a given month.

5.1 Malaria Policies Toward Travelers

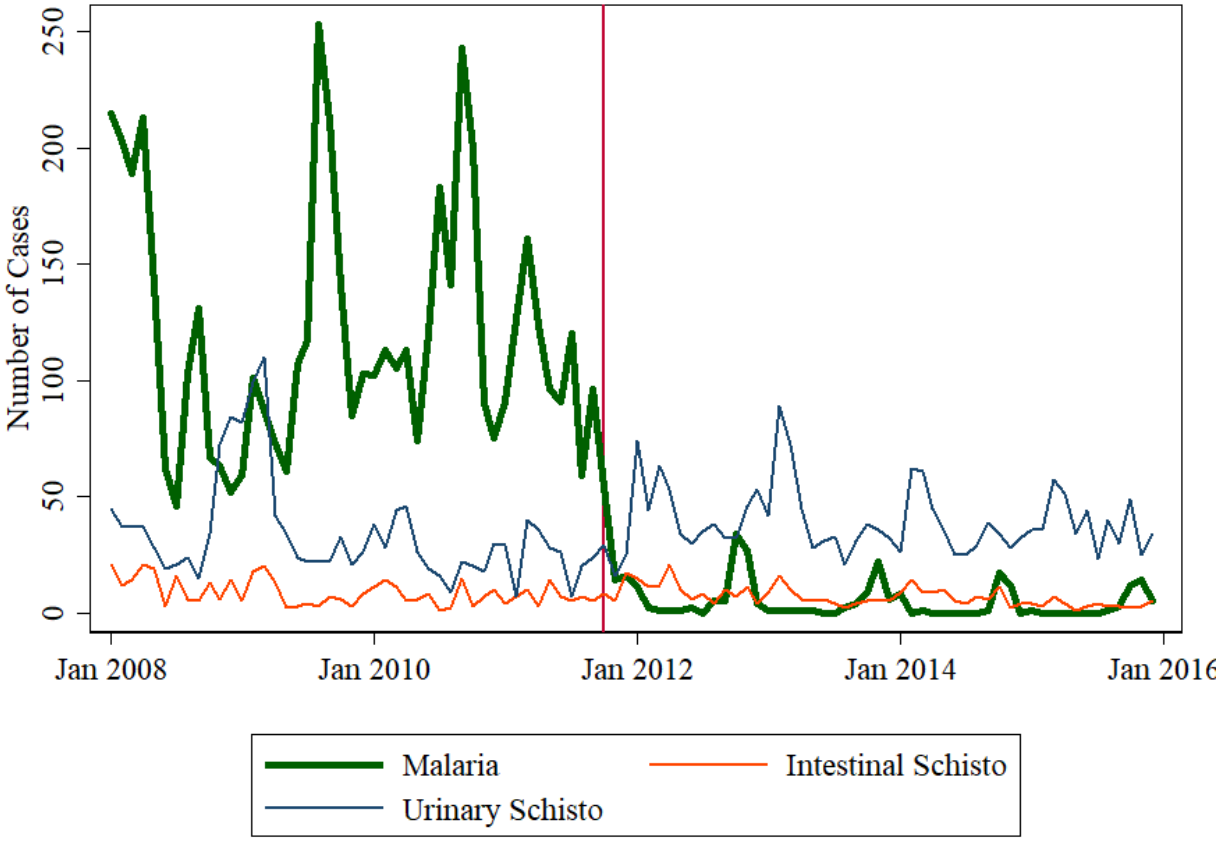
The NGO PATH has specifically focused on reducing imported cases in the district Richard Toll. They have used volunteers in the community to alert health workers to the arrival of new travelers. Health workers track down these travelers and ask to test them using an RDT, and treat those that test positive. 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; therefore, this strategy was only able to detect 5% of imported cases.[25]

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 it (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. There was a drastic decrease in cases after the implementation of this policy, with case numbers at zero or close to zero after the policy (Figure 5). Data on two types of schistosomiasis among workers at the CSS show no drop in those diseases after late 2011, demonstrating that the drop in malaria cannot be attributed to an overall improvement in the health care facility. This closely mirrors the outcomes measured by Dillon, Friedman, and Serneels (2014), who find that a policy offering testing and malaria treatment for workers at a sugarcane plantation in Nigeria leads to a 10% increase in earnings due to increased labor supply and productivity.

A strategy to decrease malaria cases that 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

[25]Data on imported cases is based on surveys conducted in Richard Toll.

Figure 5: 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 for malaria and treated those that tested positive. Data was provided by the CSS.

every household in a village, checking for individuals with symptoms. 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 policy could be applied to areas that receive travelers during the weeks when the most expected infections enter.

Finally, just as mobile phones can be used to measure aggregate movement patterns to improve targeting, they could also be used more directly for the targeting. Mobile phones are always linked to a tower if they are turned on; therefore, the provider knows as soon as an individual changes towers. An innocuous policy could be to allow users to opt into an information program that sends a targeted text message as soon as someone that has opted into the program changes location from a high malaria tower to a low malaria tower. The message could recommend and provide an incentive to get tested at the closest clinic for free. Text messaging has been used by the Ministry of Health in Senegal for providing information on diabetes, and other studies have found that SMS technology can be effective in changing behavior (Senegal Ministry of Health 2016, Aker and Ksoll 2019, Ksoll et al. 2014, Fischer et al. 2016). Mobile phones could be used for more controversial policies such as targeting mandatory quarantine. There are important ethical concerns in how new technology is used, and the severity and risk of the disease may affect the strategies that are considered.

5.2 Targeting Simulations

The policies in the previous section serve as examples, but I focus the simulations on how a policy should be targeted, rather than the particular policy. If there were no targeting and instead the policy were to test every single traveler coming into the 5 pre-elimination districts that are the focus here, it would mean testing 6,956,197 travelers in 2013 based on the scaled mobile phone data. If each traveler were successfully tested so that all travel cases of malaria were found, treated, and secondary infections were prevented, it would mean 663 cases treated or prevented, representing around 44% of total cases in this area. Even just taking the cost of an RDT to test each person ($0.50), without adding any other costs

associated with such an intervention, it amounts to around $5,300 per case. Compared to a typical benchmark for cost-effectiveness of $150, targeting all travelers would be 35 times as costly. Therefore, if a policy were to target travelers, only a subset should be targeted.

There are two ways that a policy can be targeted: (1) it can target certain destination areas for travelers or (2) it can target travelers from specific origin areas. For both cases there are two sets of costs: (1) fixed cost, which consists of training community health workers, health post nurses, and district health supervisors in investigation of travel cases as well as costs for weekly electronic data transmission; and (2) variable cost, which are the costs associated with investigating and treating travelers.[26] I use the Fiscal Year 2015 Malaria Operational Plan for Senegal for the policy simulations. Based on this plan, $400,000 was allocated for case investigation and targeting individuals in districts with incidence less than 5/1000. Given there were 19 such districts in 2015, I assigned the funding proportional to district population in order to assign an amount to the five districts that are part of the simulation. This leads to $127,850 dedicated to the five low malaria districts studied here. I assume an even split between fixed and variable costs. To calculate the variable cost per person, I use the information available from Richard Toll where 3,609 people were targeted in 2015. Variable funding for that district of $10,286 leads to a per person cost of $2.85.

The first type of targeting focuses on destination areas. Given some set of resources R meant to target travelers, those resources can either be distributed across all health facilities or they can be concentrated in a smaller set of facilities in certain areas and in certain months. The most naive method for targeting would be to randomly select health facility areas and months and target all travelers in the areas/months randomly selected. This is in effect equivalent to not targeting. In practice this would never be done since at minimum policymakers know when malaria is most prevalent and would not target travelers during months the country is effectively malaria-free. Therefore, a second strategy would be to target randomly certain health areas during the months of high malaria prevalence first (August

[26]For simplicity, I assume the variable cost scales proportionally with the number of travelers.

to December) and then during the low malaria prevalence months. With the information available at hand, policymakers could do even better in targeting by using information from the previous year. They know which health facility areas/months had the highest level of malaria in the previous year and can target health facility areas in months according to their ordering the previous year.

The cell phone data allow for even more effective targeting. The piece of information currently not available to policymakers is the number of travelers entering an area from any given other district in the country. The cell phone data provide this information and combined with the incidence from 2012, it makes it possible to estimate which health facility areas are most at risk from travelers in certain months. This can be especially effective because travel choices are not random. There are often high movement corridors between certain communities in a country. Therefore, it is likely that there are pockets that will be more affected by travelers coming from high malaria areas than other parts of the district. By focusing on the pockets of areas most affected by travelers from high malaria settings, limited resources can be spent more effectively training health professionals and tracking travelers only in these areas.

Figure 6 demonstrates the full cost curves, from a strategy where only one health district area-month is treated, all the way up to all health district area-months being treated. The benefit at each point is calculated based on the earlier model to calculate the total primary and secondary cases attributed to travelers. I assume that all travelers are targeted, but only 94% of them agree to take the test and contribute to the benefit from targeting.[27] Depending on the resources available to the government, it is possible to determine for any given budget how many cases would be treated or averted depending on the targeting strategy chosen. Looking across the entire distribution in the top panel, the targeted strategy based on information from mobile phones is consistently the most cost- effective one. Across the whole distribution, the cell phone data based targeting policy performs 11.15% better on

[27] This is based on data from Richard Toll where 3,386 out of 3,609 targeted travelers agreed to be tested.

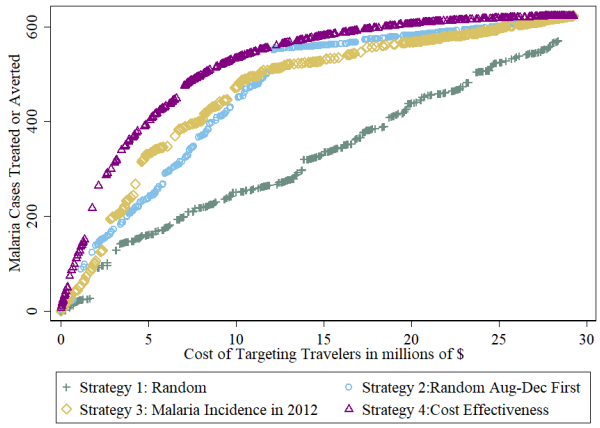
average compared to the next best policy of using information on incidence from the year before. Zooming in on the part of the distribution up to $400,000, the amount spent on this type of program, the cell phone data targeted policy performs over 300% better on average (Panel b, Figure 6).

The second type of targeting focuses on travelers from particular origin areas. This is more difficult in terms of implementation because it requires the ability to identify travelers coming from specific areas, but has the potential of being more cost-effective since resources are focused on the highest risk travelers. I base the targeting strategies on the district the individuals are coming from and the month in which they are traveling. 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. I assume again that 94% of targeted travelers are tested based on the experience in Richard Toll. Four strategies are again compared: (1) the implausible one of the government randomly choosing district-months (non-targeting); (2) the government choosing district-months randomly within the malarial season and then during the rest of the year; (3) the government using monthly malaria incidence in the prior year to order the district-months from those with the highest to the lowest incidence; and (4) a cost-benefit value is calculated for each district-month based on the number of travelers and the variable cost and uses incidence from the previous year to calculate the total impact of the imported cases (the benefit).

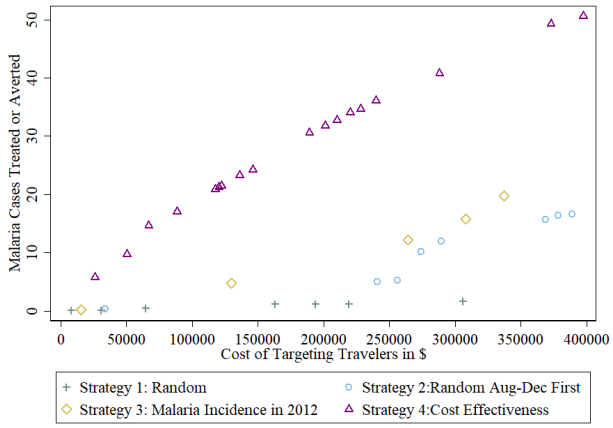
Panel a of Figure 7 shows the cost curves for these four strategies zoomed in on the relevant budget of $400,000. Note that the cost starts at $63,928, which is the total fixed cost for all the facilities in the five districts, since it is assumed that training and preparations are done across all health facility areas but only particular travelers are targeted across these areas. Using the cell phone data to inform targeting leads to higher cost effectiveness than the next best strategy based on the government using incidence from the previous year to target travelers from specific locations. In particular, at a budget of $400,000 it performs 27% better on average. Panel b compares targeting of specific travelers entering all five

Figure 6: Targeting Areas: Cost and Benefit Under Different Strategies

(a) Full Cost Curve



(b) Cost Curve Zoomed in on Less than $400,000



Notes: The panels show four different strategies for targeting travelers. Each symbol represents a health facility area-month. Targeting a specific health facility area-month means targeting all travelers entering that health facility area in that month. The strategies lay out which health facility area-months are targeted first. The cost is calculated based on a variable cost of $2.85 per traveler and a fixed cost of $63,928 split proportionally between health facility areas based on population and number of facilities. The benefit is based on the parameters of the model to calculate the number of primary and secondary cases generated by travelers from each district in each month and summed for all travelers in a given health facility area month. It is assumed that only 94% of those targeted are successfully tested.

low malaria districts (type 2) to targeting all travelers that enter only specific health post catchment areas in those five districts (type 1), using the cell phone data for both. Targeting specific travelers rather than only areas is 257% more effective with a budget of $400,000.

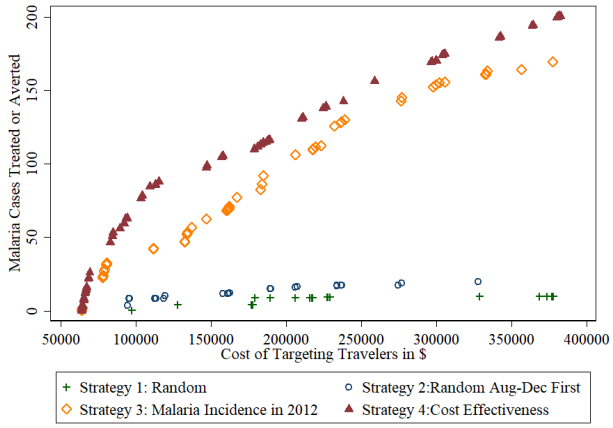
The previous scenarios consider either targeting particular travelers or targeting particular areas, but it is possible to combine the two types of targeting to target particular travelers going to particular areas in certain months. This would not be possible without the cell phone data, which provides extremely granular information on spatial movement of individuals across time so that we know how many travelers enter a specific area in a given month and this can be used to calculate the cost of targeting each of these travelers and the benefit of quickly identifying and treating any cases they may have brought through travel. Figure 8 shows the cost curve in this scenario where travelers from a particular district to a particular health facility area are targeted in a specific month.[28] This curve is compared to the previous curves of just targeting travelers, just targeting areas, and also the best strategy without using cell phone data of targeting travelers based on incidence in 2012. Targeting both travelers and areas leads to 52% better performance on average compared to the non-cell phone data strategy when focusing on a budget of under $400,000. Using cell phone data for targeting both travelers and areas compared to just targeting travelers is 19% more effective on average with a budget under $400,000.

There are two important limitations in conducting this type of targeting. The first is related to potential risk of targeting areas based on movement information from a previous year, given that population movement patterns may change drastically from one year to the next. Other research that has used cell phone data for several years in Namibia finds very consistent short term movement patterns across three years (Milusheva et al. 2017, Wesolowski, Erbach-Schoenberg, et al. 2017). Additionally, if short term movement matrices from cell phone data were made available to policymakers on an ongoing basis, it would be possible to adjust targeting in real time as information becomes available.

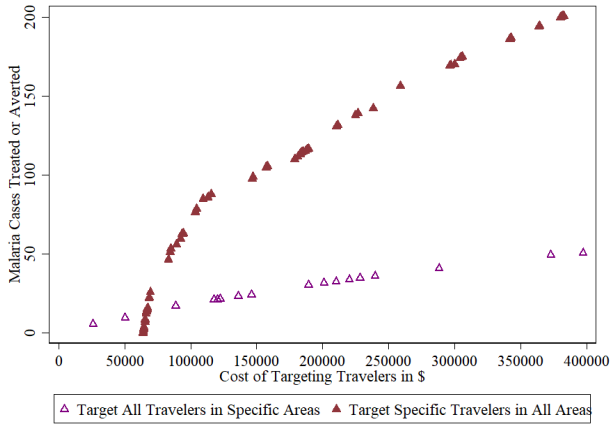
[28]The fixed cost of an area is added when the first set of travelers in a month is treated in that area.

Figure 7: Targeting Travelers: Cost and Benefit Under Different Strategies

(a) Cost Curve Zoomed in on Less than $400,000

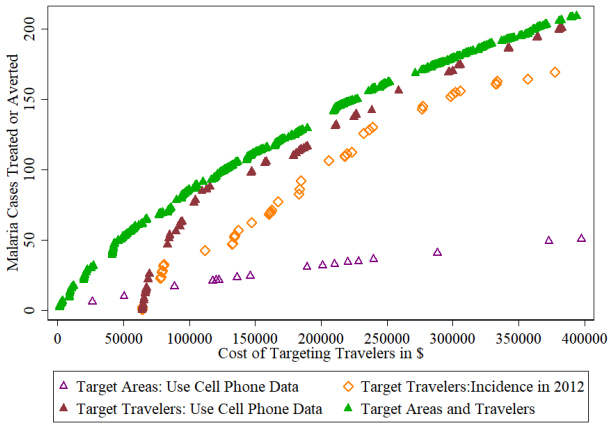


(b) Comparing Targeting Travelers and Targeting Areas



Notes: Panel (a) shows 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. Panel (b) compares targeting of specific travelers entering all five low malaria districts to targeting all travelers that enter only specific health post catchment areas in those five districts, using the cell phone data for both. The cost is calculated based on a variable cost of $2.85 per traveler and a fixed cost of $63,928 split proportionally between health facility areas based on population and number of facilities. The benefit is based on the parameters of the model to calculate the number of primary and secondary cases generated by travelers from each district in each month and summed for all travelers in a given health facility area month. It is assumed that only 94% of those targeted are successfully tested.

Figure 8: Targeting Both Travelers and Areas Compared to Different Strategies



Notes: A scenario where travelers from a particular district to a particular health facility area are targeted in a specific month is compared to previous scenarios of targeting particular health facility areas, targeting travelers from particular districts, and the best-case scenario if no cell phone data is available of targeting travelers based on the incidence of districts in the previous year.

The second limitation relates to representativeness and who may be missed through these targeting strategies (Blumenstock 2018). The movement patterns of the lowest income are likely missing from these data due to lack of a mobile phone. Therefore, this type of targeting may miss marginalized areas that may experience importation of malaria from very low income groups that are not captured in the movement patterns. If mobile phone data targeting strategies are implemented by policymakers, it could lead to marginalized pockets of malaria in the elimination zones due to lack of targeting to the areas receiving the lowest income travelers. Luckily, elimination zones have in place surveillance systems at health facilities that track malaria cases that come to the facility. Thus, it will be possible to analyze the data for outlier health facilities that do not experience a decrease in malaria after the targeting strategies are implemented.

Up to now, high malaria districts from where cases are imported have not been discussed. If malaria were reduced significantly in those districts, it would automatically reduce importation. The assumption is that while strategic targeting is done in the low-malaria zones, in the high-malaria districts, a package of interventions aimed at reducing the burden of the disease is maintained. This is in line with the WHO strategy for malaria elimination.

6 Robustness Checks

I do several robustness checks to test the main specification (Table 3). Column 2 uses an estimate of imported incidence without weighting the mobility data to be representative of the full population. This assumes that the only movement in the country is the movement in the Sonatel data, which would be an underestimate. Weighting the data 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 real effect should also lie between these upper and lower bounds of 2.1 and 1.7. In addition to weighting the movement, Column 3 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 the main specification (Column 1). Utilization is relatively uniform throughout the country; therefore, incorporating utilization does not significantly change the results. I rerun the specification using Conley standard errors that account for spatial autocorrelation and serial autocorrelation over time (Column 4) (Hsiang 2010).[29] The results remain significant.

In Column 5, I use net imported incidence, subtracting expected infected travelers leaving the health post catchment area. Both the coefficients on imported and lagged imported do not change significantly and they remain significant. Results from additional robustness checks are available in Appendix B. These include (1) using cases rather than incidence; (2) adjusting the population that incidence is based on by the number of people entering and leaving the area each month; (3) loosening the assumption that the susceptible population is equal to 1 and scaling lagged incidence by a calculated susceptible population; and (4) restricting the movements included in calculating imported incidence by removing any movement where the SIM card had no calls or texts for more than two weeks before or after.

One limitation of the mobile phone data is that it only includes mobility within Senegal and cannot incorporate international migration. Immigrants coming in from high malaria

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

Table 3: Robustness Checks

	(1) Baseline Model	(2) Unweighted	(3) Scaled by Utilization	(4) Spatial and Autocorrelation SE	(5) Net Imported
Imported Incidence	1.094***	1.563***	1.294***	1.230***	1.217***
	(0.357)	(0.509)	(0.449)	(0.326)	(0.384)
Lag Incidence	0.563***	0.563***	0.567***	0.330***	0.561***
	(0.129)	(0.129)	(0.128)	(0.0543)	(0.130)
Rain in cm	-0.000654	-0.000654	-0.00119	0.00640	-0.000852
	(0.00790)	(0.00790)	(0.0206)	(0.00863)	(0.00788)
Lag Rain in cm	0.0269	0.0269	0.0703	0.0292*	0.0258
	(0.0181)	(0.0181)	(0.0466)	(0.0167)	(0.0180)
Lag 2 Rain in cm	0.0319**	0.0319**	0.0812**	0.0380**	0.0318**
	(0.0156)	(0.0156)	(0.0398)	(0.0150)	(0.0156)
Constant	-0.0688**	-0.0688**	-0.181**		-0.0663**
	(0.0290)	(0.0290)	(0.0756)		(0.0288)
Health Post x Month Obs	432	432	432	432	432

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

Notes: Augmented version of the Arellano-Bond Generalized Method of Moments estimator used in all specifications except for Column 4. In Column 4, I use Conley standard errors, which account for spatial autocorrelation and serial autocorrelation over time, but do not cluster at the health post catchment area level. In Column 2, I do not weight movement observations to scale up to the full population of Senegal. In Column 3, I scale the expected imported incidence and total incidence by the utilization of the region based on DHS data. In Column 5, I calculate net imported incidence by subtracting out the expected cases leaving a health facility area divided by the population of the area.

countries as well as emigrants returning and visiting family from high malaria settings could both impact malaria incidence. There are detailed case data available from Richard Toll district where each case was investigated and travel information was included on the infected individual based on survey data. Out of 161 cases, there are 9 cases where the traveler was from outside of Senegal, or only 5.59% of cases. The small impact of international migration likely arises from the fact that while some of Senegal's neighbors such as Mali have a higher malaria incidence (89 cases per 1000), the only international border near the pre-elimination districts studied here is Mauritania, which has a lower incidence rate than northern Senegal of only 0.4 cases per 1000. Nevertheless, future work could try to incorporate the impact of international travelers.

7 Conclusion

The paper quantifies the negative externality of population movement on disease incidence and reversing gains in elimination in Senegal, and it proposes a cost-effective targeting strategy. This type of study is made possible by new big data collected by telecommunications companies making the measurement of short term movement possible and the initiative taken by the Ministry of Health to collect surveillance data on a monthly level for each health post. These data made it possible to study how short term movement into low-malaria areas increases incidence. The study finds that for each imported case of malaria per 1000, there are around 1.7 cases of malaria per 1000 reported at health post areas in the pre-elimination districts. Using these findings, I calculate that implementing a policy in certain destination areas and targeting travelers from the most cost-effective districts during particular months results in the largest drop in cases (over 50% as many cases treated or averted as compared to the next best strategy). To our knowledge, this is the first study to evaluate the cost effectiveness of different types of targeting schemes for malaria directed at travelers.

While the work here focuses on malaria, it is possible to implement this type of model

for other infectious diseases such as influenza, Ebola, Zika or Coronavirus. As cell phone usage has become prevalent throughout the developing world and cell phone providers are beginning to understand how the data they collect can be used by policymakers to implement better policies, measuring short term movement becomes easier. The collection and integration of high frequency data on infectious diseases then makes it possible to study these models that help policymakers better target interventions, and then document their impact. This could help countries lower the disease burden from existing infectious diseases and prevent epidemics of new diseases, leading to beneficial economic and social consequences.

References

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

Adhvaryu, Achyuta (2014). "Learning, misallocation, and technology adoption: evidence from new malaria therapy in Tanzania". *The Review of economic studies* 81.4, pp. 1331–1365.

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

Aker, Jenny C and Christopher Ksoll (2019). "Call Me Educated: Evidence from a Mobile Phone Experiment in Niger". *Economics of Education Review*.

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

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

Apouey, Bénédicte, Gabriel Picone, and Joshua Wilde (2018). "The Economics of Malaria Prevention". *Oxford Research Encyclopedia of Economics and Finance*.

Armand, Alex et al. (2017). "Do public health interventions crowd out private health investments? Malaria control policies in Eritrea". *Labour Economics* 45, pp. 107–115.

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

Barofsky, Jeremy, Tobenna D Anekwe, and Claire Chase (2015). "Malaria eradication and economic outcomes in sub-Saharan Africa: evidence from Uganda". *Journal of health economics* 44, pp. 118–136.

Bjorkegren, D and Darrell Grissen (2018). "Behavior revealed in mobile phone usage predicts credit repayment". *World Bank Economic Review. Accepted.*

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

Blumenstock, Joshua (2016). "Fighting poverty with data". *Science* 353.6301, pp. 753–754.

— (2018). *Don't forget people in the use of big data for development.*

Blumenstock, Joshua, Gabriel Cadamuro, and Robert On (2015). "Predicting poverty and wealth from mobile phone metadata". *Science* 350.6264, pp. 1073–1076.

Blumenstock, Joshua, Nathan Eagle, and Marcel Fafchamps (2016). "Airtime transfers and mobile communications: Evidence in the aftermath of natural disasters". *Journal of Development Economics* 120, pp. 157–181.

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

Buckee, Caroline O, Andrew J Tatem, and Jessica Metcalf (2017). "Seasonal population movements and the surveillance and control of infectious diseases". *Trends in parasitology* 33.1, pp. 10–20.

Casey, Nicholas (2016). "Hard Times in Venezuela Breed Malaria as Desperate Flock to Mines". *The New York Times* August 15, 2016.

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

Chang, Hsiao-Han et al. (2018). "The geography of malaria elimination in Bangladesh: combining data layers to estimate the spatial spread of parasites". *BioRxiv*, p. 421578.

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

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

Cohen, Jessica and Pascaline Dupas (2010). "Free distribution or cost-sharing? Evidence from a randomized malaria prevention experiment". *The Quarterly Journal of Economics*, pp. 1–45.

Cohen, Jessica, Pascaline Dupas, and Simone Schaner (2015). "Price subsidies, diagnostic tests, and targeting of malaria treatment: evidence from a randomized controlled trial". *The American Economic Review* 105.2, pp. 609–645.

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

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

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

Currie, Janet and Tom Vogl (2013). "Early-life health and adult circumstance in developing countries". *Annual Review of Economics* 5.1, pp. 1–36.

Cutler, David et al. (2010). "Early-life malaria exposure and adult outcomes: Evidence from malaria eradication in India". *American Economic Journal: Applied Economics* 2.2, pp. 72–94.

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

Dillon, Andrew, Jed Friedman, and Pieter Serneels (2014). *Health information, treatment, and worker productivity: Experimental evidence from malaria testing and treatment among Nigerian sugarcane cutters*. The World Bank.

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

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

Dupas, Pascaline (2014). "Short-run subsidies and long-run adoption of new health products: Evidence from a field experiment". *Econometrica* 82.1, pp. 197–228.

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.

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

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

Fischer, Henry H et al. (2016). "Text message support for weight loss in patients with prediabetes: a randomized clinical trial". *Diabetes Care* 39.8, pp. 1364–1370.

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

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

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

Hong, Sok Chul (2011). "Malaria and economic productivity: a longitudinal analysis of the American case". *The Journal of Economic History* 71.3, pp. 654–671.

Hong, Sok Chul (2013). "Malaria: An early indicator of later disease and work level". *Journal of health economics* 32.3, pp. 612–632.

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

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

Ihantamalala, Felana Angella et al. (2018). "Estimating sources and sinks of malaria parasites in Madagascar". *Nature communications* 9.1, p. 3897.

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

Ksoll, Christopher et al. (2014). "Learning without Teachers? A Randomized Experiment of a Mobile Phone-Based Adult Education Program in Los Angeles". *Center for Global Development Working Paper* 368.

Laxminarayan, Ramanan et al. (2010). "Should new antimalarial drugs be subsidized?" *Journal of health economics* 29.3, pp. 445–456.

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

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

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

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

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

Lucas, Adrienne M (2010). "Malaria eradication and educational attainment: evidence from Paraguay and Sri Lanka". *American Economic Journal: Applied Economics* 2.2, pp. 46–71.

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

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

Milusheva, Sveta et al. (2017). "Understanding the Relationship between Short and Long Term Mobility". *AFD Research Paper Series* 2017.69.

Montalvo, Jose G and Marta Reynal-Querol (2007). "Fighting against malaria: prevent wars while waiting for the "miraculous" vaccine". *The Review of Economics and Statistics* 89.1, pp. 165–177.

Montjoye, Yves-Alexandre de et al. (2014). "D4D-Senegal: The Second Mobile Phone Data for Development Challenge".

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

Nickell, Stephen (1981). "Biases in dynamic models with fixed effects". *Econometrica: Journal of the Econometric Society*, pp. 1417–1426.

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

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

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

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

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

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

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

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

Senegal Ministry of Health (2016). *M-Diabete: Le Mobile au Service de la Lutte contre le Diabete.* URL: `http://www.mdiabete.sante.gouv.sn/`.

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

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

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

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

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

Tarozzi, Alessandro et al. (2014). "Micro-loans, insecticide-treated bednets, and malaria: evidence from a randomized controlled trial in Orissa, India". *American Economic Review* 104.7, pp. 1909–41.

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

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

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

Venkataramani, Atheendar S (2012). "Early life exposure to malaria and cognition in adulthood: evidence from Mexico". *Journal of health economics* 31.5, pp. 767–780.

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

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

Wesolowski, Amy, Elisabeth zu Erbach-Schoenberg, et al. (2017). "Multinational patterns of seasonal asymmetry in human movement influence infectious disease dynamics". *Nature communications* 8.1, p. 2069.

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

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

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

WHO (2020a). *Coronavirus disease 2019 (COVID-19) Situation Report 52*. URL: https://www.who.int/docs/default-source/coronaviruse/20200312-sitrep-52-covid-19.pdf?sfvrsn=e2bfc9c0_2.

— (2020b). *Ebola virus disease*. URL: https://www.who.int/news-room/fact-sheets/detail/ebola-virus-disease.

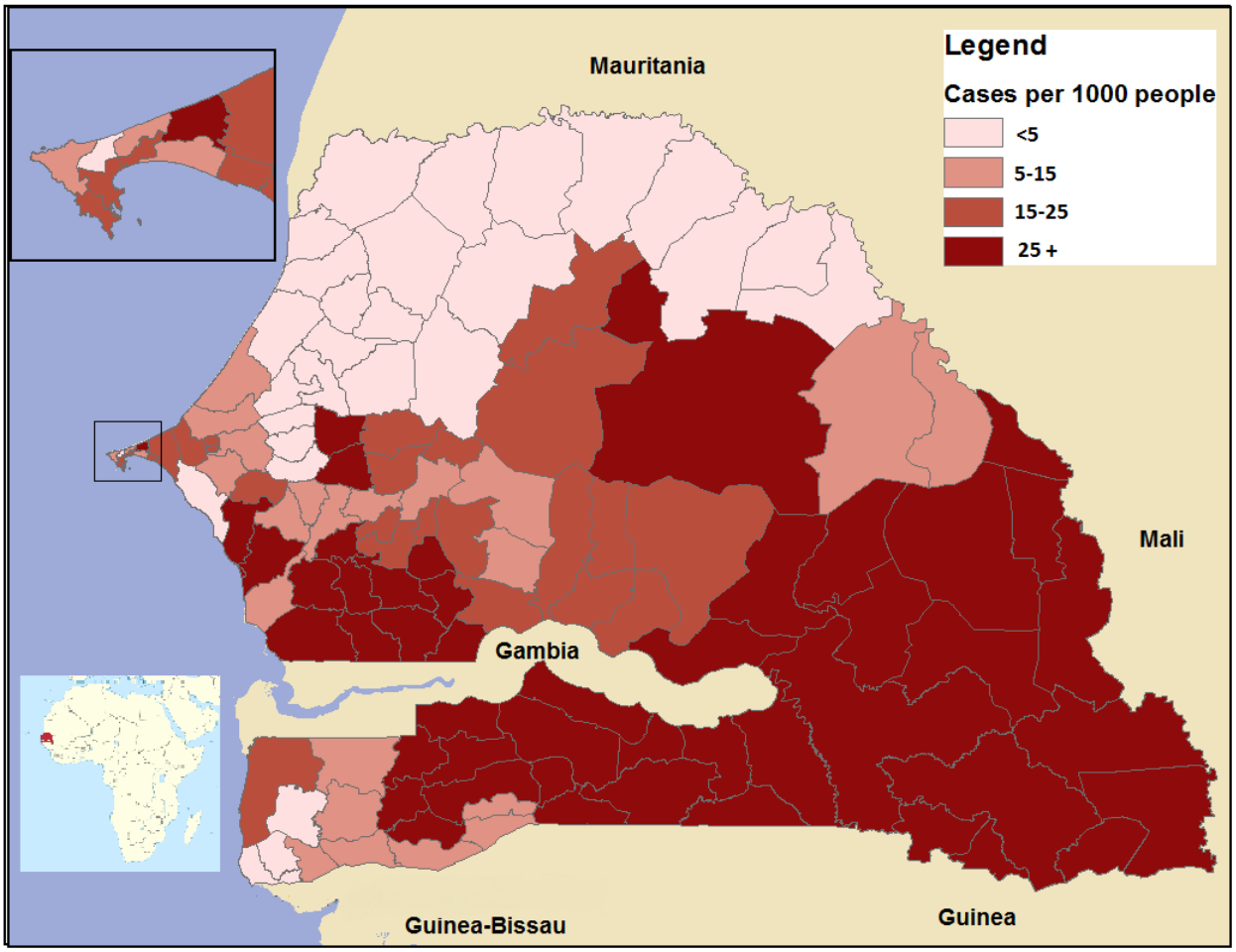
Wiser, Mark (2010). *Protozoa and human disease*. Garland Science.

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

World Malaria Report (2019). Tech. rep. World Health Organization.

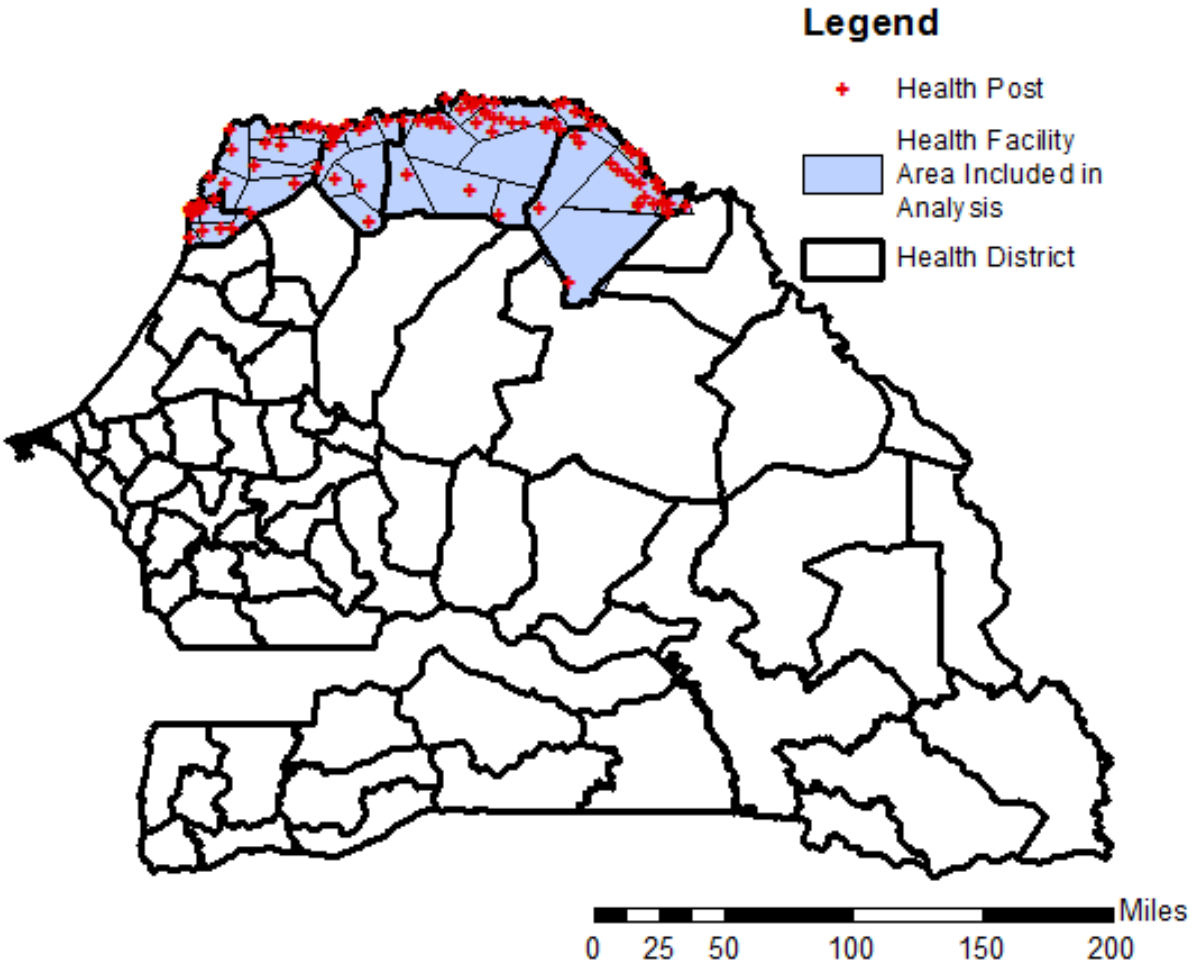
A Additional Tables and Figures

Figure A.1: Annual Malaria Incidence in Senegal in 2013



Notes: Data come from tested and confirmed cases at the health district level compiled by the National Malaria Control Program.

Figure A.2: Senegal Health Districts and Location of Health Posts in the North Used in the Analysis



Notes: The five very low malaria health districts used in the analysis are subdivided into health post catchment areas that group health posts and mobile phone towers together.

Table A1: Using Different Rainfall Specifications

	(1) Baseline	(2) No Rain Lags	(3) Only 1 Month Lagged Rain	(4) Only 2 Months Lagged Rain	(5) Quadratic Functional Form of Rain	(6) Logged Rain
Imported Incidence	1.094***	1.313***	1.338***	1.203***	1.087***	1.284***
	(0.357)	(0.416)	(0.407)	(0.354)	(0.414)	(0.452)
Lag Incidence	0.563***	0.597***	0.598***	0.563***	0.510***	0.658***
	(0.129)	(0.125)	(0.123)	(0.135)	(0.138)	(0.133)
Rain in cm	-0.000654	-0.00846			-0.000416	
	(0.00790)	(0.00845)			(0.0315)	
Lag Rain in cm	0.0269		0.0170		-0.0602	
	(0.0181)		(0.0169)		(0.0518)	
Lag 2 Rain in cm	0.0319**			0.0228	-0.0395	
	(0.0156)			(0.0142)	(0.0469)	
(Rain in cm)^2					-0.000200	
					(0.00174)	
Lag (Rain in cm)^2					0.00429	
					(0.00299)	
Lag 2 (Rain in cm)^2					0.00369	
					(0.00274)	
Log Rain in cm						-0.0117
						(0.0125)
Lag Log Rain in cm						-0.0232
						(0.0204)
Lag 2 Log Rain in cm						0.0224**
						(0.0111)
Constant	-0.0688**	-0.0532**	-0.0686**	-0.0555**	-0.00706	-0.0379
	(0.0290)	(0.0237)	(0.0301)	(0.0222)	(0.0269)	(0.0308)
Health Post x Month Obs	432	432	432	432	432	283
Month FE	Yes	Yes	Yes	Yes	Yes	Yes

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

Notes: Augmented version of the Arellano-Bond Generalized Method of Moments estimator used in all specifications.

Table A2: Robustness Checks

	(1) Baseline	(2) Cases	(3) Adjusted Pop	(4) Scale by Susceptible Pop	(5) Restrict Phone Data
Imported Incidence/Cases	1.094***	1.077***	1.070***	1.091***	1.495**
	(0.357)	(0.191)	(0.376)	(0.349)	(0.601)
Lag Incidence/Lag Cases	0.563***	0.684***	0.619***	0.625***	0.575***
	(0.129)	(0.157)	(0.0716)	(0.0867)	(0.128)
Rain in cm	-0.000654	-0.273	0.000636	-0.000471	-0.00140
	(0.00790)	(0.191)	(0.00822)	(0.00811)	(0.00792)
Lag Rain in cm	0.0269	0.188	0.0243	0.0268	0.0262
	(0.0181)	(0.639)	(0.0178)	(0.0183)	(0.0181)
Lag 2 Rain in cm	0.0319**	0.860*	0.0322**	0.0306**	0.0315**
	(0.0156)	(0.507)	(0.0148)	(0.0150)	(0.0152)
Constant	-0.0688**	-1.579	-0.415	-0.142	-0.0719**
	(0.0290)	(1.199)	(0.264)	(0.164)	(0.0298)
Month FE	Yes	Yes	Yes	Yes	Yes
Health Post x Month Obs	432	396	432	432	432

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

Notes: Augmented version of the Arellano-Bond Generalized Method of Moments estimator used in all specifications. In Column 2, I use cases rather than cases per 1000 people. In Column 3, the population of the health post catchment area is adjusted monthly by the number of people entering and exiting based on the mobile phone data scaled by a factor of .7 to account for the population not represented by these data. In Column 4, lag incidence is scaled by the percent of population susceptible to malaria. In Column 5, movements are restricted to those where the SIM card had a call or a text within 14 days of when the movement is counted.

B Model of Malaria

Due to the two-host system, malaria is modeled using two differential equations to describe the dynamics of infected humans and infected mosquitoes. These were first modeled by R. Ross (1910) and then 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, although here, I will present it in discrete time. The state variables are y_{it}, the fraction of mosquitoes infected in location i at time t and x_{it}, the fraction of humans infected in location i at time t.

In the model extension here that includes the impact of population movement, cases can either be generated locally in location i or imported from any other location j into i, $\mathcal{I}_i$. As modeled by Torres-Sorando and Rodriguez (1997), imported cases enter the model linearly. Local cases in location i are generated based on the susceptible population and several biological parameters. The susceptible population, S_{it-w}, is the fraction of the population that is susceptible to malaria at time $t-w$, where w is the total parasite incubation period. The other parameters are the transmission efficiencies from infected mosquitoes to humans and humans to mosquitoes, b_{it-w} and c_{it}, the number of bites on humans per mosquito, a_{it-w} and the ratio of mosquitoes to humans, m_{it-w}. The change in the number of infected humans and mosquitoes is described by:

$$y_{it} - y_{it-w} = a_{it-w}c_{it-w}x_{it-w}(e^{-\mu_{it-w}\tau_{it-w}} - y_{it-w}) - \mu_{it-w}y_{it-w} \quad (3)$$

$$x_{it} - x_{it-w} = m_{it-w}a_{it-w}b_{it-w}y_{it-w}S_{it-w} + \mathcal{I}_{it} - r_i x_{it-w} \quad (4)$$

where μ_{it} is the mortality rate of mosquitoes and τ_{it} is the incubation period from the time a mosquito becomes infected until it is infectious. The parameter r_i is the recovery rate of humans. I make four assumptions arising from the epidemiology literature based on modeling the system in a low malaria setting:

1. The total parasite incubation time is one month based on the malaria cycle.

2. The mosquito population is at the steady state since mosquito populations have a relatively rapid turnover [30]

3. All malaria cases in month t are treated immediately and recover in month t.[31]

4. Based on D. L. Smith and McKenzie (2004), when the proportion of infected humans is small, the number of infectious bites received per day by a human (known as the entomological inoculation rate, taking the form $EIR_{it} = \frac{m_{it}a_{it}^2c_{it}e^{-\mu_{it}\tau_{it}}x_{it}}{\mu_{it}+a_{it}c_{it}x_{it}}$) can be approximated by $c_{it}C_{it}x_{it}$, where C_{it} is the expected number of humans infected per infected human per day, assuming perfect transmission efficiency ($b_{it} = c_{it} = 1$), known as the vectorial capacity. The assumption applies because the analysis is conducted

[30] According to the CDC, adult female mosquitoes, which spread malaria, do not live more than 1-2 weeks in nature (Center for Disease Control 2015).

[31] The incidence data in the analysis is based on diagnosed cases, which are provided with free antimalarial treatment upon diagnosis. The literature shows that within a few days of treatment the majority of parasites are eliminated (Nosten and Nicholas J White 2007, N. White 1997).

in a low malaria setting. Figure B.1 provides evidence that the assumption holds for these data.

Based on assumption 1, I set $w = 1$. Based on assumption 2, I solve equation 3 for the quasi-equilibrium proportion of infectious mosquitoes as has been done in D. L. Smith and McKenzie 2004 and Ruktanonchai et al. 2016:

$$y_{it-1} = \frac{a_{it-1}c_{it-1}x_{it-1}e^{-\mu_{it-1}\tau_{it-1}}}{\mu_{it-1} + a_{it-1}c_{it-1}x_{it-1}} \tag{5}$$

Assumption 3 implies the recovery rate, r_i, is equal to one since all infected individuals recover within the same month. This allows me to focus on new cases of malaria in month t. Since immunity does not develop in low incidence areas, anyone who is not currently infected is susceptible to the disease, which implies that the susceptible population is 1 if everyone recovers. Rewriting equation 4 to incorporate the implications of assumptions 1-3:

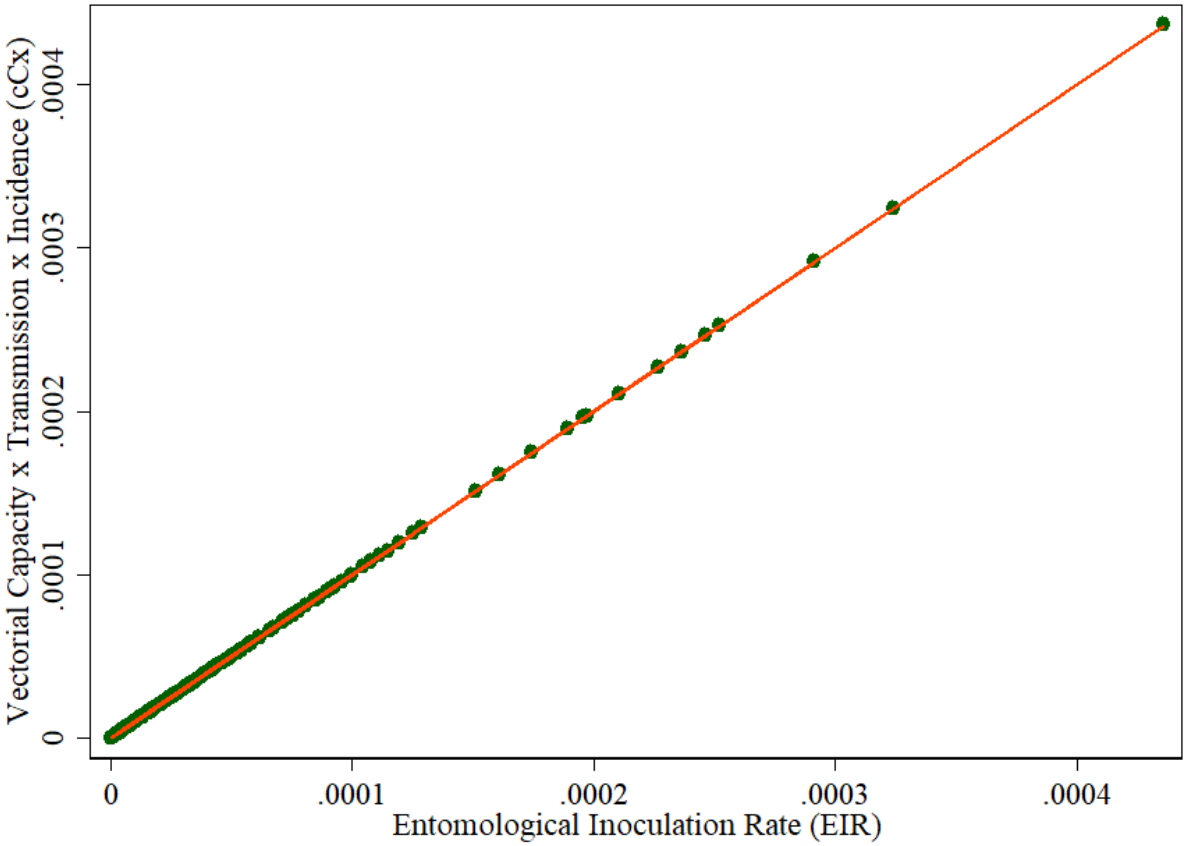
$$x_{it} - x_{it-1} = \frac{a_{it-1}^2 b_{it-1}c_{it-1}m_{it-1}e^{-\mu_{it-1}\tau_{it-1}}x_{it-1}}{\mu_{it-1} + a_{it-1}c_{it-1}x_{it-1}} + \mathcal{I}_{it} - x_{it-1} \tag{6}$$

Based on assumption 4, equation 6 can be rewritten as:

$$\begin{aligned} x_{it} &= b_{it-1}EIR_{it-1} + \mathcal{I}_{it} \\ &= b_{it-1}c_{it-1}C_{it-1}x_{it-1} + \mathcal{I}_{it} \end{aligned}$$

Assumption 4 is important because by rewriting EIR_{it-1} as $c_{it-1}C_{it-1}x_{it-1}$, 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. It is then possible to estimate this model using OLS as described in the main text of the paper.

Figure B.1: Testing the Approximation of the Entomological Innoculation Rate in Assumption 4



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 4, in low malaria districts it should be possible to approximate EIR with cCx. In the scatter plot, this means the points should be along the 45 degree line, which is shown in red. The graph shows that for the districts analyzed here, the assumption holds. The EIR and cCx were estimated using values from the literature for the various biological malaria parameters. Based on Gething et al. (2011), average transmission from infected humans to mosquitoes, c, is 0.161. The incubation period, τ, is 9 days (Killeen, A. Ross, and T. Smith 2006) and the bites on humans per mosquito, a, is 0.3 (Ruktanonchai et al. 2016). Average mosquito lifespan of 12.5 is used.