This is a repository copy of *Malaria Suitability, Urbanization and Subnational Development in Sub-Saharan Africa*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/165194/

Version: Accepted Version

Article:

Flückiger, Matthias orcid.org/0000-0002-2242-0220 and Ludwig, Markus (2020) Malaria Suitability, Urbanization and Subnational Development in Sub-Saharan Africa. Journal of Urban Economics. 103279. ISSN 0094-1190

https://doi.org/10.1016/j.jue.2020.103279

Reuse

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Malaria Suitability, Urbanization and Subnational Development in Sub-Saharan Africa*†

Matthias Flückiger‡ Markus Ludwig§

September 3, 2020

Abstract

Using subnational data, we document that the climatic suitability for malaria falciparum transmission constitutes a first-nature characteristic that influences today's spatial distribution of urbanization and socioeconomic development in Sub-Saharan Africa. Both, levels of urbanization and development are lower in regions that exhibit a high malaria transmission potential. Evidence further indicates that the settlement behavior of the European colonizers plays an important role in explaining why urban areas are concentrated in low risk areas. Throughout, we rely on an exclusively climate-based measure of malaria falciparum transmission intensity that is independent of local prevalence rates for identification. Robustness of estimates to inclusion of climatic suitability indices for further tropical diseases, null results in placebo tests and reproduction of findings outside of Africa support the validity of our identification strategy.

JEL CLASSIFICATION: I15, R12, O18, N57

KEYWORDS: Malaria, Urbanization, Economic Development, Sub-Saharan Africa

*We are grateful for valuable comments received from the editor (Stuart Rosenthal), two anonymous referees, Davide Cantoni, Hartmut Egger, Stephan Heblich, Erik Hornung, Frank Krysiak, Jürg H. Sommer, Uwe Sunde, and Mark Westcott for valuable suggestions and comments. We also thank conference participants at the 2015 CEA conference in Toronto, the 2015 World Congress of the Econometric Society in Montréal, the 2016 RES conference in Brighton, the 2016 CSAE conference in Oxford, the 2016 AEHN conference in Brighton, and the 2017 Development Economics and Policy Conference in Göttingen. Furthermore, we thank seminar participants at the University of Basel, University of Bayreuth, Queen's University Belfast, TU Dresden, and Ludwig-Maximilians University Munich.

†An earlier version of this paper was circulated under the title 'Malaria Suitability and (Post-)Colonial Urbanization in Sub-Saharan Africa'.

‡University of York. Email: `matthias.flueckiger@york.ac.uk`

§Technische Universität Braunschweig. Email: `markus.ludwig@tu-braunschweig.de`

1 Introduction

Plasmodium falciparum malaria has for millennia posed a risk to health in Sub-Saharan Africa (Loy et al., 2017).[1] In this study, we assess to what extent differences in the local climatic potential for Plasmodium falciparum malaria have generated within-country disparities in urbanization and economic development. To identify effects, we construct a measure of the local climatic malaria transmission potential that is specifically tailored to the Plasmodium falciparum parasite and is exogenous with respect to local socioeconomic conditions. This measure is based on a biological model that describes mosquito and parasite development as non-monotonic functions of air temperature. Parametrization exclusively relies on experimental data obtained from laboratory studies (Mordecai et al., 2013). The resulting measure captures the local, time-invariant climatic suitability for malaria falciparum transmission. It is best be interpreted as a locational fundamental akin to other geographical features.[2] This first nature characteristic not only influences current-day outcomes via direct effects on contemporaneous health, but also through its health-related effects in the past that influenced long-term determinants of economic development. The latter may include culture, institutions or locational choices of where to establish settlements. When interpreting results, it is important to keep in mind that they represent reduced-form effects, i.e., the net effect of all channels through which the malaria intensity measure affects outcomes.[3]

To our knowledge, we are the first to use an exogenous measure for malaria transmission suitability to analyze its effect on current-day urbanization and development. Studies that investigate related questions typically employ malaria-intensity indices that are (partly) based on observed prevalence. These measures are unsuited for the purposes of our analysis due to potential endogeneity. For illustration, take an uninhabited (and undeveloped) area. In this region the malaria prevalence is zero, even when the climatic potential for malaria transmission is very high. However, the area may be deserted precisely owing to the potentially detrimental effect of malaria. Conversely, a higher level of economic development implies that more resources are available to control malaria, thus reducing the number of infections. Due to the existence of these types of reverse causality, the

[1] In this paper, we will use the terms 'malaria' and 'Plasmodium falciparum malaria' interchangeably. In both instances we refer to the latter type of malaria.

[2] Note that the climatic malaria potential represents a locational fundamental only as long as the malaria parasite and its host (the mosquito) are present. Absent either component, the measure does not represent a negative local characteristic. See Sections 2 and 4.2 as well as Appendix B for more details.

[3] This also implies that we cannot use the malaria environment as an instrument for contemporaneous malaria prevalence. The existence of a multitude of channels violates the exclusion restriction.

use of a prevalence-based malaria measure would (likely) bias our estimates. Similarly, the widely used global malaria stability index developed in Kiszewski et al. (2004) is potentially endogenous as it depends on the actual distribution of vectors and the observed human biting rates of mosquitos. Both these aspects can be influenced by human economic activity (see e.g., Killeen et al. (2001), Seyoum et al. (2002), Guerra et al. (2006), Ngwa (2006), Vittor et al. (2006), Smith and Ruktanonchai (2010), or Mutuku et al. (2011)).

In addition to introducing an exogenous climatic measure for Plasmodium falciparum malaria transmission intensity, our paper makes two main contributions to the economic literature. First, we systematically document that the malaria environment constitutes a determinant of today's spatial distribution of urbanization and economic activity in Sub-Saharan Africa. To this end, we divide Sub-Saharan Africa into 5,841 grid cells of 0.5×0.5 degree. For each cell we determine the local degree of malaria suitability and then assess its effect on urbanization and economic activity using a cross-sectional OLS regression approach. Throughout, we control for temperature, a variety of additional exogenous physical characteristics as well as country fixed effects. Consequently, we identify effects of the malaria environment by exploiting the residual within-country variation between the non-monotonic, malaria-falciparum-specific function of air temperature and generic—malaria-unrelated—temperature effects. The resulting estimates show that the local climatic malaria potential strongly deters urbanization. Urban population decreases by 0.117 standard deviations when moving from a region characterized by an average malaria transmission suitability to an area in which the malaria potential is one-standard deviation higher. This translates into a reduction of 31,678 individuals. The deterrent effect on economic activity, proxied by night-time light intensity, is similar in magnitude.

As a second contribution, we show that the local malaria potential also influences the spatial distribution of socioeconomic development. For this part of the analysis, we draw on nationally representative cross-sectional survey data collected in the Demographic and Health Surveys (DHS) Program. Our sample covers 28 countries and encompasses 522,538 individuals aged 15–49 who reside in 287,858 distinct households. Using this data, we document that persons living in areas with a high climatic malaria transmission suitability accumulate less human capital (measured by educational attainment), are less likely to be employed in the more productive non-agricultural sectors, and accumulate less wealth. Again, the implied magnitude of the point estimates is economically meaningful. Household wealth, for example, decreases by 16 percent (evaluated at the sample mean) when the malaria potential increases by one standard deviation.

Taken together, our results document that the climatic malaria transmission potential constitutes a locational fundamental that deters socioeconomic development and urbanization. In the final step of our analysis, we investigate the relevance of one potential channel that links the malaria environment to urbanization: the locational choices of the European colonizers. Specifically, we show that European economic centers were less likely to emerge in high malaria risk regions. Many of these centers have since evolved into large cities. This, in turn, helps explain why today's spatial distribution of urbanization is influenced by the local malaria environment.

The key identifying assumption underlying our analysis is that malaria suitability index specifically captures the climatic malaria transmission potential rather than some generic effect of temperature. We support the validity of this assumption by exploiting the fact that the penetration of the Plasmodium falciparum parasite in Latin America varied over time and space (Webb (2008, p. 73, p. 85 ff.)). We show that variation in the malaria transmission potential does not predict urbanization or development in regions where the parasite was absent. In areas with a stable presence of the parasite, on the other hand, the malaria environment constitutes a locational fundamental that strongly deters these two outcomes. Further evidence for the malaria-specificity of our results is provided by the fact that our estimates remain stable when we control for temperature suitability indices developed for alternative tropical diseases, including Yellow and Dengue fever as well as Animal Trypanosomiasis. To document the stability of our estimates more generally, we conduct a battery of robustness checks. These include extending the set of control variables, employing alternative measures for urbanization, modifying the malaria suitability index, and using different standard error clustering approaches.

Our paper is related to different strands of literatures. An influential body of work looks at the relationship between the malaria environment and current-day economic development. Several cross-country studies document negative correlations (e.g., Gallup et al. (1999), Acemoglu et al. (2001) or Gallup and Sachs (2001)). A number of papers that use sub-national data to identify determinants of economic development also takes into account measures of malaria exposure (e.g., Henderson et al. (2012), Michalopoulos and Papaioannou (2013), Michalopoulos and Papaioannou (2014), Alsan (2015), Henderson et al. (2018)). While the majority of these studies find a negative relationship between malaria and current-day development outcomes, the point estimates are potentially biased due to the use of prevalence-based malaria intensity indices. Furthermore, no study specifically focuses on assessing the effect of the malaria environment on the spatial distribution of development. Rather, the malaria index is treated as a control variable and its impact is not discussed in detail.

Linked to our investigation are further papers that analyze the effects of the malaria environment on development in Sub-Saharan Africa during the pre-colonial period (Depetris-Chauvin and Weil, 2018; Easterly and Levine, 2016). For this earlier period, the studies do not find any negative effects. This implies that the negative relationship only emerged during the 20th century, i.e., the colonial and post-colonial era. We provide additional empirical support for this view and suggest the malaria-influenced settlement pattern of the European colonizers as a possible explanation for the emergence of the negative effect. In this regard, our results also connect to the discussion about the effects of European colonial activity on current-day economic development (e.g., Acemoglu et al. (2001); Putterman and Weil (2010); Easterly and Levine (2016); Ali et al. (2018)) as well as the papers that document a strong degree of persistence in the spatial distribution of urbanization (e.g., Redding et al. (2010); Michaels and Rauch (2018); Jedwab et al. (2017); Jedwab and Moradi (2016)).

By identifying the malaria environment as an important locational fundamental, our study also speaks to the literature that analyzes the effect of geography on urbanization (e.g., Davis and Weinstein (2002); Rappaport and Sachs (2003); Bosker et al. (2007); Rosenthal and Strange (2008); Saiz (2010); Miguel and Roland (2011); Motamed et al. (2014); Henderson et al. (2017, 2018)) or economic development (e.g., Acemoglu et al. (2003); Rappaport and Sachs (2003); Acemoglu et al. (2005); Dell (2010); Dell et al. (2012); Nunn and Puga (2012); Alsan (2015); Flückiger and Ludwig (2017)).

Finally, this paper connects to the large body of research that analyzes the effects of malaria on individual-level outcomes (e.g., Bleakley (2010); Cutler et al. (2010); Lucas (2010, 2013)). These studies generally investigate the effects of malaria eradication campaigns on proximate determinants of development, contemporaneous health in particular. Our estimates, on the other hand, capture variation in socioeconomic outcomes that are caused by differences in the local, climatically determined, potential for malaria transmission. Apart from exerting a direct effect on contemporaneous health this locational characteristic can affect economic well-being via its influence on the evolution of more fundamental determinants of development, such as local institutions or culture. Crucially, malaria-induced differences in these deep fundamentals persist even if malaria is eradicated.

The remainder of the paper is structured as follows: In the next section, we introduce our measure of climatic malaria transmission suitability in detail and outline the methodology employed in the empirical analysis. In Section 3, we present the data along with a descriptive analysis thereof. The regression results are discussed in Section 4, after which potential channels are investigated. We conclude with Section 6.

2 Empirical Strategy

In this section, we first discuss the inherent endogeneity between the spatial distribution of human population and malaria prevalence in more detail. We then outline our strategy to circumvent this issue and describe the regression methodology employed in the empirical analysis.

Endogeneity between the Spatial Distribution of Human Population and Malaria Prevalence

In order for Plasmodium falciparum malaria to be prevalent in a region, three necessary conditions have to be met: (1) temperature conditions must allow for vector and parasite development, (2) human population density must be sufficiently high, and (3) both, the falciparum parasite and a vector, i.e., a mosquito, capable of transmitting the parasite must be present. The last two conditions are, in contrast to the first one, endogenous with respect to the spatial distribution of human population. Condition (2) implies, inter alia, that malaria prevalence is zero in uninhabited regions. Similarly, the spatial distribution of vectors (condition (3)) is endogenous with respect to past and present human activity and population densities. Local agricultural practices or housing conditions, for example, influence which vector establishes itself as the dominant one (e.g., Amerasinghe et al. (1991), Tuno et al. (2005), McNeill (2010, p.65ff)). Furthermore, human interventions, such as eradication campaigns, can influence the regional diffusion of vector or parasite. The existence of such two-way relationships imply that identification of malaria-related effects is not possible by relying on measures that incorporate any components that are based on, or influenced by, observed malaria prevalence, vector distribution or human population densities.

Climate-Based Malaria Transmission Suitability Model

Our measure of local malaria transmission is based on the Basic Reproductive Number (R_0), i.e., the number of malaria cases that arise from one case introduced into a population of susceptible hosts. This epidemiological metric is commonly used to capture the malaria transmission risk. We base our measure on the R_0 metric developed by Mordecai et al. (2013), who model local transmission intensity as a multitude of air-temperature-sensitive functions that reflect vector and parasite development. Two important characteristics motivate the use of the Mordecai et al. (2013) model: First, parametrization and functional forms of the individual components of the metric are based solely on experimental data from laboratory studies that analyze the temperature-dependent parasite and vector development. The results of these studies are—unlike values derived from field studies—independent of local differences in human population densities and human activity (Mordecai et al., 2013, p. 25). Second, the Mordecai et al. (2013) model incorporates all aspects of vector and parasite development as functions of air temperature. Alternative metrics typically only model a limited number factors that influence R_0 as temperature-dependent functions (e.g., Weiss et al. (2014)).[4]

Formally, our malaria transmission intensity model consists of three multiplicative components, each representing non-linear functions of air temperature.[5] The metric of our model is given by:

$$R_0(T) = \left(\frac{M(T)\, \mathrm{e}^{-\mu(T)\, E(T)}}{\mu(T)} \right)^{1/2}, \tag{1}$$

where T represents air temperature, $M(T)$ the mosquito density, $\mu(T)$ the mosquito mortality rate, and $E(T)$ the extrinsic incubation period of the Plasmodium falciparum parasite in the mosquito. The fact that $M(T)$, i.e. the mosquito density, is itself a function of temperature-dependent vector development implies that the measure is fully characterized by the thermal physiology of the vector and the parasite (Mordecai et al., 2013, p. 25). The key difference between the original Mordecai et al. (2013) model and our modified version presented in Eq.(1), is that the original measure incorporates additional components which are potentially endogenous with respect to local socioeconomic conditions (e.g., via processes of co-evolution). Specifically, these are the human biting rate of the anopheles, the proportion of the bites of infective mosquitoes that infect susceptible humans as well as the proportion of bites of susceptible mosquitoes on infectious humans that infect mosquitoes. To avoid any endogeneity issues, we do not model these components as temperature-dependent functions.

Panel (a) of Figure 1 depicts the relationship between air temperature and the malaria

[4] There is a large number of models that derive R_0 based on a mix of laboratory and field studies (e.g., Martens et al. (1995); Ermert et al. (2011); Lunde et al. (2013a)).

[5] The functions are either quadratic functions or left-skewed unimodal functions of air temperature. The functions and their parameter values are reported in Table 2 in Mordecai et al. (2013).

transmission intensity metric, $R_0(T)$. The optimal temperature for malaria transmission is around 25°C. At temperatures below 17°C and above 34°C, the climatic potential for malaria transmission is zero.

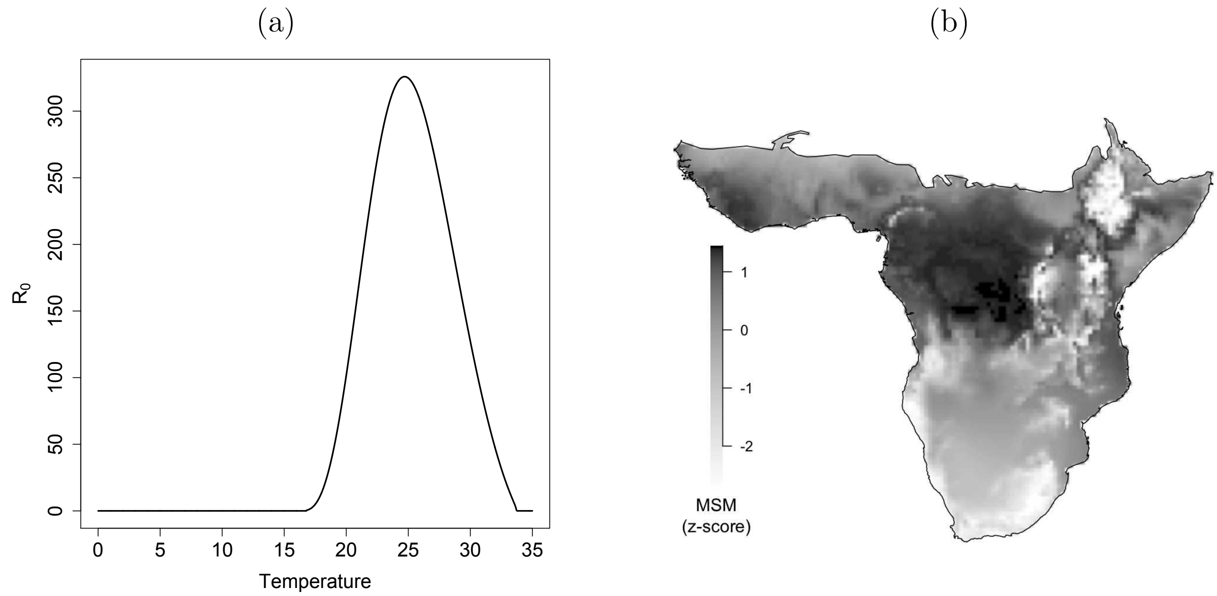

Figure 1: Panel (a) represents R_0 as a function of air temperature. Panel (b) depicts Geographical distribution of the MSM (z-score) in Sub-Saharan Africa. The darker the shading, the higher the climatic suitability for Plasmodium falciparum malaria transmission.

We construct a time-invariant measure of the local malaria potential by computing the mean of the monthly malaria transmission potential, R_0, defined in Eq.(1), over the time period 1901–1925.[6] Formally, our time-invariant malaria transmission suitability measure (MSM) can be expressed as:

$$MSM = \frac{1}{300} \sum_{y=1901}^{Y=1925} \sum_{m=1}^{M=12} R_0(T_{m,y}). \tag{2}$$

The MSM is computed as the sum of the individual monthly malaria transmission intensity values ($R_0(T_{m,y})$) in month m of year y over the time span 1901–1925. This value is then divided by the number of months in the 25-year time period, i.e., 300. The geographical distribution the resulting malaria transmission suitability measure is depicted in Panel

[6]We focus on a period during which urbanization had not reached any significant levels in Sub-Saharan Africa in order to avoid possible endogeneity issues related to the possibility that the extent of urbanization influences temperatures (e.g., Arnfield (2003)). However, as shown in Tables C.5–C.6, our result are robust to changing the time interval used to compute the MSM.

(b) of Figure 1. The highest malaria risk is observed for Central Africa. Further clearly discernible are elevated and therefore cooler areas where malaria transmission suitability is considerably lower compared to neighboring regions.

Regression Methodology

We conduct our empirical analysis at two levels of aggregation: The grid cell level and the individual level. Below, we illustrate our empirical approach at the grid cell level. The methodology translates directly to the individual level.

The following OLS regression equation is used to analyze the effect of malaria suitability on urbanization and economic activity:

$$y_{g,c} = \theta\, MSM_{g,c} + \boldsymbol{\beta}'\, \mathbf{X}_{g,c} + \tau_c + \varepsilon_{g,c}. \tag{3}$$

Current-day urbanization (or economic activity) in grid cell g located in country c is represented by $y_{g,c}$. Our measure for the grid-cell-level climatic malaria suitability is symbolized by $MSM_{g,c}$. Vector $\mathbf{X}_{g,c}$ contains exogenous geographical and climatic variables that are commonly associated with urbanization and economic development. In our baseline specification this set of controls encompasses temperature, precipitation, relative humidity, their squared values, the respective first order interaction terms as well as longitude, (absolute) latitude and a tropics dummy. These variables account for the fact that climatic conditions can directly influence urbanization and economic development. Consequently, the identifying variation of the malaria suitability index is given by the difference between the vector and parasite development functions of temperature constituting the MSM measure and the climate variables, particularly temperature and temperature squared. This variation is generated to some part by the kinks in the R_0-function (Figure 1, panel (A)) but also by averaging subannual variation in the non-linear suitability metric (R_0) over time, as described in Eq.(2).

In all regressions we account for country-specific fixed effects (τ_c). These allow for the possibility that the ability to cope with a given level of climatic malaria suitability varies with country-specific characteristics, such as the level of development or quality of national institutions.[7] Throughout, we standardize variables (i.e. build z-scores) to facilitate

[7] Our results remain unchanged if we additionally include ethnicity dummies in our regression setup. Based on the map of Murdock (1959), each grid cell is assigned the ethnicity with the largest overlapping territory. This assignment procedure produces 632 ethnicity dummies. Regression outputs are available upon request.

comparison of point estimates and cluster standard errors at 2×2 degree grid cells.

3 Data and Descriptive Analysis

Data

For our analysis at the grid-cell level, we divide mainland Sub-Saharan Africa into 5,841 grids cells of 0.5×0.5 degrees and focus on two outcome variables: urban population numbers and economic activity. The former are taken from Jedwab and Moradi (2016), who provide population estimates for localities with more than 10,000 inhabitants covering the period 1890–2010.[8] In our main analysis, we use the population estimates for the year 2010, the latest year available. Our proxy for local economic activity, is based on nighttime luminosity data provided by the Defense Meteorological Satellite Program-Optical Line Scanner (DMSP-OLS) sensor (see Henderson et al. (2012)). In keeping with the population data, we construct our dependent variable for the year 2010 and add up the light-intensity indices of the individual 1×1 km pixels that fall into a given 0.5×0.5 degree grid cell.

For each grid cell, we compute the time-invariant malaria suitability measure, as well as mean temperature, mean precipitation, and mean relative humidity by taking the respective average over the period 1901–1925 in analogy to Eq.(2). The monthly climate data are provided by the Climatic Research Unit of the University of East Anglia (CRU TS version 3.22). Additionally, we determine the caloric suitability index (Galor and Özak, 2015, 2016), mean elevation (NOAA, 2009) and distance to the coastline for each cell. We further generate an indicator variable that takes the value one if a cell is intersected by a waterway, and, zero otherwise. This information is extracted from `www.naturalearthdata.com`. Using the shapefiles available from `www.gadm.org` we identify which country each cell belongs to.[9,]

[8] Jedwab and Moradi (2016)'s dataset is based on information reported in *Africapolis I: West Africa* and *Africapolis II: Central & Eastern Africa*. *Africapolis*, in turn, uses a wide array of sources—including population censuses (reports and gazetteers), administrative counts, demographic surveys and electoral counts—to create a consistent database for African cities in 33 countries. Using the same type of sources, Jedwab and Moradi (2016) expand the data to an additional six countries. In some instances, population numbers are linearly interpolated within decades to harmonize population estimates across data sources and years (see Web Appendix page A.4 and Table 12 Jedwab and Moradi (2016) for more detail). As a result, information is available for the whole of Sub-Saharan Africa except Lesotho, South Africa and Swaziland.

[9] When a grid cell overlaps multiple countries, it is assigned to the country that occupies the largest share of the cell.

In the individual-level analysis, we draw on data from the Demographic and Health Surveys (DHS) Program. For our purposes, nationally representative data (men and women aged 15–49) are available for 28 Sub-Saharan African countries. When multiple survey waves exist, we use the most recent one.[10] The waves span the period 2006–2016. As dependent variables, we use educational attainment, a household wealth index, non-agricultural occupation, as well as a dummy variable indicating whether a household is located in an urban area.[11,12] Details regarding the construction of these variables are presented in Table A.1. To link the survey data to the malaria environment and other location-specific characteristics we use geocoded information on the households' residence provided by the DHS. Our final dataset encompasses 522,538 individuals from 287,858 households.

Descriptive Analysis[13]

To verify that the spatial variation in the MSM indeed captures the potential for malaria transmission, we employ regression model (3) and document that the MSM influences various grid-cell level prevalence measures. In Table 1 of column (1) we regress average estimated Plasmodium falciparum malaria prevalence over the period 2000–2010 (taken from Bhatt et al. (2015)) on the MSM. The dependent variable in column (2) is observed malaria prevalence derived from the DHS.[14] In addition to current-day prevalence, we look at the frequency of sickle haemoglobin alleles in the general population in column (3).[15] This genetic mutation offers protection against malaria, but is associated with otherwise considerably lower life expectancy (Hedrick, 2011). Due to the existence of this trade-off, the frequency of the mutation in populations today is representative of malaria transmission intensity in the past (Depetris-Chauvin and Weil, 2018). As displayed in Table 1, the

[10] Table A.2 provides a table containing details regarding the countries and survey waves included in our analysis.

[11] The DHS wealth index is a principal component index based on ownership dummies for a fixed set of durable goods.

[12] An important caveat pertains to the definition of urban residence in the DHS data. Urban residence is coded according to the classifications of the National Statistical Offices. In some cases levels of non-agricultural employment form part of the administrative definition of urban areas. This can lead to an mechanical correlation between urban residence and non-agricultural employment in the DHS data (see International Labour Organization, 2015). This should be taken into account when interpreting the results presented in Table 3.

[13] Summary statistics of the key variables are provided in Tables A.4–A.5.

[14] We only include grid cells for which survey information on malaria prevalence is available. The sample size is consequently reduced. We obtain very similar results if we run the regression at the individual level, with an indicator for being infected with Plasmodium falciparum malaria at the time of interview as the dependent variable.

[15] Data produced by Piel et al. (2013) and available at the grid cell level for the whole of Africa.

MSM exerts a positive and highly significant effect on all measures of prevalence. This documents that our climatic suitability measure successfully captures the local malaria falciparum transmission risk, both in the past and present.

Table 1: Malaria Suitability Measure and Prevalence

	Prevalence (SD) Bhatt et al. (2015) (1)	Prevalence (SD) DHS (2)	Sickle Cell (SD) Piel et al. (2013) (3)
MSM (SD)	0.090***	0.214***	0.085**
	(0.030)	(0.056)	(0.042)
Climate controls	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes
Observations	5,841	1,676	5,841
R-squared	0.899	0.532	0.825

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at 2×2 degree grid cells and reported in parentheses. All regressions control for land surface area of the grid cells. Climate controls include temperature, precipitation and relative humidity, the squared terms of these variables and their first-order interactions as well as longitude, (absolute) latitude and a Tropics dummy. 'Prevalence (SD) Bhatt et al. (2015)' represents the average grid-cell level Plasmodium falciparum infection prevalence over the period 2000–2010, as provided by Bhatt et al. (2015). 'Prevalence (SD) DHS' is computed as the (sample-weighted) average Plasmodium falciparum prevalence of the population residing within a given grid cell reported in the AIS, DHS and MIS surveys that include a Malaria Rapid Diagnostic Test Results module (see Table A.3). 'Sickle Cell (SD)' is the predicted frequency of sickle haemoglobin alleles in the general population taken from Piel et al. (2013).

4 Malaria Suitability, Urbanization and Development

In this section, we first document that the climatic malaria environment influences the spatial distribution of urbanization as well as socioeconomic development in Sub-Saharan Africa. We then discuss the validity of our identification strategy and robustness of results.

4.1 Main Results

Malaria Suitability, Urbanization and Economic Activity at the Grid-Cell Level

We start by regressing urban population on our malaria suitability index, a basic set of climate controls as well as country-fixed effects in Panel I of Table 2. The result, presented in column (1), documents that the malaria environment exerts a statistically highly significant deterrent effect: A one-standard deviation increase in the MSM reduces grid-cell urbanization by 0.135 standard deviations. In column (2), we augment the regression setup to include the Tsetse fly suitability index. As shown in Alsan (2015), economic development in regions characterized by a hostile Tsetse environment was lower in the pre-colonial

Table 2: Intensive Margin: Malaria Suitability and Current-Day Grid-Cell-Level Urban Population

	Current Day Urban Population		
	(1)	(2)	(3
Panel I: overall (total urban population)			
MSM (SD)	-0.135***	-0.127**	-0.11
	(0.051)	(0.051)	(0.0
Observations	5,370	5,370	5,3
R-squared	0.051	0.051	0.0
Panel II: extensive margin (urban population yes/no)			
MSM (SD)	-0.163***	-0.150***	-0.12
	(0.043)	(0.046)	(0.0
Observations	5,370	5,370	5,3
R-squared	0.216	0.217	0.2
Panel III: intensive margin (total urban population conditional on any urban population)			
MSM (SD)	-0.168**	-0.162**	-0.15
	(0.078)	(0.072)	(0.0
Observations	1,429	1,429	1,4
R-squared	0.074	0.074	0.0
Country fixed effects	Yes	Yes	Ye
Climate controls	Yes	Yes	Ye
TseTse control	No	Yes	Ye
Geography controls	No	No	Ye

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at 2×2 degree grid cells and reported in parentheses. Each coefficient reported in the table represents the point estimate for MSM (SD) obtained from running a separate version of regression Eq.(3). All regressions control for the land surface area of the grid cell. Climate controls include temperature, precipitation and relative humidity, the squared terms of these variables and their first-order interactions as well as longitude, (absolute) latitude and a Tropics dummy. TseTse is the TseTse suitability index developed in Alsan (2015). Geography controls include distance to coast, elevation, waterway indicator, and caloric suitability.

era, with negative effects persisting up to the present day. A natural concern is that our malaria suitability measure partially captures these Tsetse effects. Reassuringly, however, our point estimate remains very similar compared to column (1). This is also the case when we control for additional geographical factors, such as distance to coastline, elevation or local caloric suitability in column (3). The implied economic magnitude of the point estimates is substantial. The coefficient of –0.117 in column (3), for example, translates into a reduction in urban population of 31,748 individuals. The effect of the MSM is also sizeable compared to other exogenous geographic variables. For instance, it is almost twice as large as the (standardized) effect of distance to coastline.

In columns (4)–(6) of Panel I, we exchange urban population with night-time light intensity as the dependent variable and re-run regressions analogous to the ones presented in columns (1)–(3). The effect of the MSM on economic activity is again negative and statistically highly significant. Night-time light intensity decreases by 0.188 standard deviations when moving from a given area to region with a one-standard deviation higher malaria suitability (column (6)).

A natural question is whether the results presented so far a driven by a particular margin of urbanization. Panels II and III of Table 2 show that this is not the case. Along both the extensive and intensive margin, we consistently find a statistically significant negative effect of malaria suitability.[16] That is, greater malaria suitability reduces (i) the probability that cities (lights) emerge and (ii) the size of cities (intensity of lights), given that cities (lights) exist.

As mentioned previously, it is important to keep in mind that our estimates capture reduced-form effects of the malaria environment. They represent the net effect of all channels through which the natural locational fundamental influences urbanization and economic activity. This includes direct health-related channels, such as increased mortality, potential reductions in labor productivity or lower human capital accumulation (Bleakley, 2010; Cutler et al., 2010). Additionally, the climatic malaria potential can also influence today's local population densities and economic development through its effect in the past. Continued, climate-induced, adverse health conditions can influence long-term determinants of growth, such as local institutions or social and cultural norms (e.g., Acemoglu et al. (2001)). Furthermore, initial malaria-influenced decisions on where to establish urban centers can have persistent effects on the spatial pattern of urbanization and economic development because of path dependence (c.f. Redding et al. (2010); Michaels and Rauch (2018); Jedwab et al. (2017); Jedwab and Moradi (2016)). While it is not possible to clearly disentangle individual mechanisms within the framework of our analysis, we explore the last channel in more detail in Section 5. Specifically, we investigate the importance of European colonizers in explaining the malaria-influenced spatial distribution of urbanization and economic activity today. First, however, we assess the effect of the climatic malaria environment on socioeconomic development.

[16]Note that the distinctions between extensive and intensive margin is somewhat arbitrary given the definition of urbanization (cut-off at 10,000 for the urban population data) and the detection threshold of satellites (night-time light data).

Malaria Suitability and Socioeconomic Development at the Individual Level

In analogy to the grid-cell level analysis, we employ the cross-sectional OLS regression model of Eq.(3). The observations are weighted using the sample weights available from the DHS. For brevity, we only report estimates obtained from regressions that include the full set of control variables (i.e., climate, TseTse, geography and country fixed effects) as well as controls for sex and age. The standard errors are clustered at the 0.5×0.5 degree grid cell level.

We start by investigating the effect of the local climatic malaria potential on educational attainment, a proxy for human capital accumulation. As documented in Table 3, columns (1)–(2), the probability of completing secondary and tertiary education declines as the malaria potential increases.[17] Evaluated at their respective sample means, the point estimates imply that a one-standard deviation increase in the MSM reduces the probability of completing secondary education by 16 percent and the likelihood of obtaining a tertiary degree by 61 percent.[18]

Table 3: Malaria Suitability and Individual and Household Level Socioeconomic Development

Dependent Variable:	Secondary Education (SD) (1)	Tertiary Education (SD) (2)	HH Wealth Index (SD) (3)	Non-Agricultural Occupation (SD) (4)	HH Urban Residence (SD) (5)
MSM (SD)	-0.121***	-0.145***	-0.332***	-0.243***	-0.399***
	(0.039)	(0.041)	(0.079)	(0.054)	(0.092)
Individual-level controls	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes
Climate controls	Yes	Yes	Yes	Yes	Yes
TseTse control	Yes	Yes	Yes	Yes	Yes
Geography controls	Yes	Yes	Yes	Yes	Yes
Observations	522,538	522,538	287,858	340,901	287,858
R-squared	0.224	0.042	0.140	0.178	0.141

Note: Standard errors are clustered at the 0.5×0.5 degree grid cell level and are reported in parentheses. All regressions control for sex and age. Climate controls include temperature, precipitation and relative humidity as well as the squared terms of these variables and their first-order interactions. TseTse is the TseTse suitability index developed in Alsan (2015). Geography controls include distance to coast, elevation, waterway indicator, (absolute) latitude, longitude, a Tropics dummy and caloric suitability. Dependent variables are constructed from DHS data. See Section 3 and Table A.1 for more information.

Column (3) looks at the effect of the MSM on a general measure of economic well-being: household wealth. When moving from a given region to an area that exhibits a one-standard deviation higher malaria transmission potential, the wealth index decreases by 0.33 standard deviations. This negative effect is consistent with the findings presented

[17] All results remain very similar if we restrict our analysis to individuals aged 25 and older.

[18] The sample means are 37 percent and 5 percent respectively.

in columns (1)–(2) as well as column (4). The latter shows that the probability of being employed in the non-agricultural sector is lower in regions with a high malaria potential.[19] Furthermore, we find that the probability of a household residing in an urban area is lower, the higher the malaria suitability is (column (5)). This last result reproduces the findings of the grid-cell level analysis: The malaria environment is an important determinant of the spatial distribution of urbanization.

Taken together, the findings presented above show that the local climatic potential for malaria transmission deters urbanization and economic development. A natural question is to what extent these differences would dissipate if malaria falciparum were to be eradicated. Empirically analyzing this question within our framework is not possible.

4.2 Threats to Identification and Robustness

The main identifying assumption in Eq.(3) is that, conditional on covariates, the MSM is unrelated to the error term and affects urbanization and individual-level outcomes only through variation in the local malaria potential. There are several potential issues pertaining to this assumption. These are discussed below.

Specificity and omitted variables

The primary threat to the validity of our analysis is that the MSM does not specifically capture the malaria risk but some other, unobserved, factor. One specific worry is that other tropical diseases that thrive under similar temperature conditions confound our findings. To investigate if this is a source of bias, we model temperature suitability for Yellow and Dengue Fever transmission following Hamlet et al. (2018) and Mordecai et al. (2017), respectively. Parametrization and functional forms of these metrics are, in analogy to the MSM, based solely on experimental data from laboratory studies.[20] As shown in Tables 4 and 5, the MSM point estimates in the grid-cell and individual-level regressions remain stable when we include these two additional disease suitability indices.[21]

While these results strongly suggest that our results are not biased due to correlation with

[19] Information regarding occupation (column (4)) is restricted to persons employed. The sample size is consequently reduced.

[20] That is, we keep components of the original metrics that are potentially endogenous to urbanization or development fixed (see Section 2) and compute the time-invariant version of these suitability measures according to Eq.(2).

[21] For brevity, we only present the results with total urban population (night-time lights) as dependent variable. Results are very similar when we alternatively look at the extensive or intensive margin.

Table 4: Controlling for Yellow and Dengue Temperature Suitability: Malaria Suitability and Current-Day Grid-Cell-Level Urban Population

	Current Day Urban Population (SD)			Night-time Lights (SD)		
	(1)	(2)	(3)	(4)	(5)	(6)
MSM (SD)	-0.135***	-0.122**	-0.123**	-0.148***	-0.155***	-0.208***
	(0.051)	(0.056)	(0.058)	(-0.055)	(0.059)	(0.062)
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Climate controls	Yes	Yes	Yes	Yes	Yes	Yes
TseTse control	No	Yes	Yes	No	Yes	Yes
Geography controls	No	No	Yes	No	No	Yes
Dengue Fever control	No	Yes	Yes	No	Yes	Yes
Yellow Fever control	No	Yes	Yes	No	Yes	Yes
Observations	5,370	5,370	5,370	5,841	5,841	5,841
R-squared	0.051	0.052	0.054	0.155	0.157	0.174

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at 2×2 degree grid cells and reported in parentheses. Each coefficient reported in the table represents the point estimate for MSM (SD) obtained from running a separate version of regression Eq.(3). All regressions control for the land surface area of the grid cell. Climate controls include temperature, precipitation and relative humidity, the squared terms of these variables and their first-order interactions as well as longitude, (absolute) latitude and a Tropics dummy. TseTse is the TseTse suitability index developed in Alsan (2015). Geography controls include distance to coast, elevation, waterway indicator, and caloric suitability. The Dengue and Yellow Fever suitability indices are taken from Mordecai et al. (2017) and Hamlet et al. (2018), respectively.

Table 5: Controlling for Yellow and Dengue Temperature Suitability: Malaria Suitability and Individual-Level Socioeconomic Development

Dependent Variable:	Secondary Education (SD) (1)	Tertiary Education (SD) (2)	HH Wealth Index (SD) (3)	Non-Agricultural Occupation (SD) (4)	HH Urban Residence (SD) (5)
MSM (SD)	-0.121***	-0.142***	-0.319***	-0.238***	-0.377***
	(0.039)	(0.042)	(0.081)	(0.055)	(0.096)
Individual-level controls	Yes	Yes	Yes	Yes	Yes
Climate controls	Yes	Yes	Yes	Yes	Yes
TseTse control	Yes	Yes	Yes	Yes	Yes
Dengue Fever control	Yes	Yes	Yes	Yes	Yes
Yellow Fever control	Yes	Yes	Yes	Yes	Yes
Geography controls	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	522,538	522,538	287,858	340,901	287,858
R-squared	0.224	0.042	0.142	0.178	0.143

Note: Standard errors are clustered at the 0.5×0.5 degree grid cell level and are reported in parentheses. All regressions control for sex and age. Climate controls include temperature, precipitation and relative humidity as well as the squared terms of these variables and their first-order interactions. TseTse is the TseTse suitability index developed in Alsan (2015). The Dengue and Yellow Fever suitability indices are taken from Mordecai et al. (2017) and Hamlet et al. (2018), respectively. Geography controls include distance to coast, elevation, waterway indicator, (absolute) latitude, longitude, a Tropics dummy and caloric suitability. Dependent variables are constructed from DHS data. See Section 3 and Table A.1 for more information.

alternative disease environments, it is still possible that the MSM captures some other climatic aspect(s) of the tropics. To substantiate our claim that this is not the case, we exploit the fact that—in contrast to Sub-Saharan Africa—the geographical diffusion of malaria falciparum was not pervasive in Latin America.[22] The Plasmodium falciparum parasite was only introduced in the 16th century with the start of the slave trade. Importantly, intensive and stable transmission of malaria falciparum was only seen in regions that saw a large influx of infectious and immune humans, i.e., African slaves (Webb, 2008, p. 79), despite vectors being present and being climatic conditions in majority of Latin America. The limited spread was mainly due to the fact that the Anopheles vectors native to the Americas are not as efficient in transmitting malaria as the African vectors, particularly the Anopheles gambiae (McNeill (2010, p. 55) and Carter and Mendis (2002)). Appendix B provides more detailed background information on the history of malaria falicparum in Latin America.

The slave-density-dependent prevalence of malaria falciparum in Latin America allows us to run both, a falsification and a validation test. In regions with low slave density—and consequently low falciparum transmission probability—the MSM should have no impact on urbanization and development if it specifically captures the malaria transmission potential. On the other hand, in regions that relied heavily on slave labor, we expect to find negative effects of the MSM, similar to the ones documented for the African continent. To test our falsification and validation hypotheses, we use data from `www.slavevoyages.org` to compute and assign the density of African slaves (expressed as slaves per square kilometer) to all countries mainland Latin America (see Table B.1).

To analyze the effect of MSM and its interplay with slave density on urbanization and economic development, we construct a grid-cell level dataset for mainland Latin America at a spatial resolution of 0.5×0.5 degrees in analogy to our main analysis.[23] The control variables included in our regression setup are identical to the ones employed in the main analysis.[24] Since the Jedwab and Moradi (2016) data only cover Sub-Saharan Africa, we alternatively use the urban population data from `citypopulation.de`.[25] We augment

[22]Today, the vast majority of the Americas is free of malaria falciparum. In areas where Plasmodium falciparum has not been eradicated, endemicity is lower than 5 percent (Gething et al., 2011). The low transmission intensity is, apart from the relatively inefficient local Anopheles vectors, the result of large scale eradication campaigns that have been implemented since the 1950s (e.g., Bleakley (2010) or Jeffery (1976)).

[23]We restrict our analysis to mainland America due to the fact the Caribbean islands possess a small surface area.

[24]The exception is that we do not include a Tsetse suitability index due to the fact that the Tsetse fly was never present in the Americas.

[25]We obtain very similar results if we construct grid-cell urban population numbers from Center for International Earth Science Information Network - CIESIN - Columbia University and International Food

regression model (3) to include an interaction term between MSM and slave density. Based on the previous discussion, we expect the MSM to exert an effect only in combination with slave density.[26]

Table 6: Malaria Suitability and Current-Day Grid-Cell-Level Urban Population in the Americas

	Urban Population (SD)		Night-time Lights (SD)	
	(1)	(2)	(3)	(4)
MSM (SD)	0.061	0.030	0.160	0.087
	(0.080)	(0.073)	(0.113)	(0.103)
MSM (SD) × slave density	-0.762***		-1.756***	
	(0.357)		(0.511)	
MSM (SD) × high slave density		-0.414**		-0.931***
		(0.192)		(0.269)
Country fixed effects	Yes	Yes	Yes	Yes
Climate controls	Yes	Yes	Yes	Yes
TseTse control	Yes	Yes	Yes	Yes
Geography controls	Yes	Yes	Yes	Yes
Observations	6,245	6,245	6,245	6,245
R-squared	0.033	0.034	0.252	0.252

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at 2×2 degree grid cells and reported in parentheses. Each coefficient reported in the table represents the point estimate for MSM (SD) obtained from running a separate version of regression Eq.(3). All regressions control for the land surface area of the grid cell. Climate controls include temperature, precipitation and relative humidity, the squared terms of these variables and their first-order interactions as well as longitude, (absolute) latitude and a Tropics dummy. Geography controls include distance to coast, elevation, waterway indicator, and caloric suitability. Brinkhoff (2015) provides census-based city-level population data for all of Latin America. The years for which population data is available vary across countries. We use the most recent census year available, which lies between 1970 and 2014.

In Table 6 we regress urban population (column (1)) and nighttime lights (column (3)) on malaria suitability. As conjectured, the uninteracted baseline effect of MSM is statistically non-significant in all regressions. This documents that absent the falciparum parasite, the climatic malaria suitability does not represent a negative local characteristic that deters urbanization. The positive sign of the uninteracted MSM coefficient suggests that, if anything, climatic malaria suitability is conducive to urbanization and economic activity in the absence of a reservoir of infectious hosts. The results supports the view that our measure specifically captures the local malaria potential and not simply some generic aspects of the tropics.

In contrast to the baseline effect, the interaction term between the MSM and slave den-

Policy Research Institute - IFPRI and The World Bank and Centro Internacional de Agricultura Tropical - CIAT (2011).

[26]Because we account for country fixed effects, direct (uninteracted) effects of slave density are absorbed by these dummies.

sity is negative and statistically significantly different from zero. This shows that with increased slave density, the climatic malaria transmission suitability becomes a negative locational fundamental that deters urbanization and economic activity. That is, combined with the presence of the Plasmodium falciparum parasite, climatic malaria suitability is an important determinant of the geographical distribution of urbanization even outside Africa. This finding provides external validation of the results presented in the main part.

In columns (2) and (4), we take a slightly different approach to analyzing the differential impact of MSM. Therein, we divide Latin America into two groups: Regions with a high density of slaves and areas with a low density of slaves.[27] We then replace the interaction term between MSM and (continuous) slave density with an interaction term between MSM and a simple indicator for high slave density. This indicator takes the value one if the cell is located within a country the belongs to the high slave density group and zero otherwise.[28] The pattern of results presented in column (2) and (4) is qualitatively equivalent to columns (1) and (3). The baseline effect of MSM is again statistically non-significant, while the interaction term is negative and statistically significant. The size of the effects for the high slave density group of countries are similar, albeit somewhat larger, compared to the point estimates obtained in the African setup. Nighttime light intensity, for example, decreases by 0.8 standard deviations as a result of a one-standard deviation increase in MSM.[29] For Africa, the corresponding coefficient estimate is –0.188.

In analogy to the grid cell analysis, we re-run the individual-level regressions for Latin America. However, DHS data are only available for Colombia and Bolivia.[30] Both these countries are part of the low slave density group, i.e., the placebo test group. For this set of countries, we therefore expect that local malaria suitability has no negative effect on the proxies for individual-level development. Panel I of Table 7 illustrates that this is the case. The MSM coefficients are small and statistically non-significant. This again suggests that in the absence of the falciparum parasite, climatic malaria suitability is not negatively—and if anything positively—associated with today's level of development.

For Brazil, by far the largest country in the high slave density group, DHS data are not

[27]As high-slave density countries classified are Brazil and the Guyanas.

[28]The use of a simple indicator rather than a continuous measure is motivated by two factors. First, the use of an indicator facilitates the comparison of coefficient sizes (Africa versus Latin America). Second, a dichotomous classification mitigates concerns related to the possibility that slave numbers are measured inaccurately.

[29]I.e., 0.087-0.931.

[30]For our analysis we use the most recent surveys (Bolivia: Phase V in 2008; Colombia: Phase VI in 2010).

Table 7: Malaria Suitability and Individual-Level Socioeconomic Development in Latin America

Panel I: Bolivia and Colombia					
Dependent Variable:	Secondary Education (SD) (1)	Tertiary Education (SD) (2)	Wealth Index (SD) (3)	Modern Sector Occupation (SD) (4)	Urban Residence (SD) (5)
MSM (SD)	0.032	0.036	0.152	-0.016	-0.010
	(0.060)	(0.049)	(0.156)	(0.060)	(0.152)
Individual-level controls	Yes	Yes	Yes	Yes	Yes
Climate controls	Yes	Yes	Yes	Yes	Yes
Geography controls	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes
Obs	72,294	72,294	49,206	72,294	49,206
R-squared	0.143	0.023	0.179	0.155	0.127

Panel I: Brazil				
Dependent Variable:	Secondary Education (SD) (1)	Tertiary Education (SD) (2)	Log Income (HH) (Log) (3)	Urban Residence (SD) (4)
MSM (SD)	-0.138***	-0.117***	-0.261***	-0.142**
	((0.050)	(0.043)	(0.077)	(0.070)
Individual-level controls	Yes	Yes	Yes	Yes
Climate controls	Yes	Yes	Yes	Yes
Geography controls	Yes	Yes	Yes	Yes
Obs.	10,770,512	10,770,512	10,770,512	10,770,512
R-squared	0.029	0.031	0.166	0.096

Note: All regressions control for sex and age. Climate controls include temperature, precipitation and relative humidity as well as the squared terms of these variables and their first-order interactions. Geography controls include distance to coast, elevation, waterway indicator, (absolute) latitude, longitude, a Tropics dummy and caloric suitability.
Panel I: Dependent variables are constructed from DHS data (women questionnaire). Standard errors are clustered at the DHS cluster level and are reported in parentheses.
Panel II: Dependent variables are constructed from the Censo Demográfico (2010). Note: Standard errors are clustered at the 0.5×0.5 degree grid cell level and are reported in parentheses.

available. We alternatively draw on data from the 2010 Brazilian census (Censo Demográfico).[31] The Brazilian data include information on the level of education, residential status (rural-urban) and annual income at the household level.

Panel II of Table 7 depicts the regression results for Brazil. Malaria suitability exerts a statistically significantly negative effect on all outcomes.[32] This result provides further

[31] The census is conducted by the Instituto Brasileiro de Geografia e Estatística (IBGE) and includes information on 72,294 individuals from 6,2 million randomly drawn households across Brazil. We thank Data Zoom (Department of Economics at PUC-Rio) for providing the codes for accessing the IBGE microdata.

[32] A potential concern is that regions characterized by a high degree of malaria transmission suitability were also areas of (disproportionately) intense plantation slavery and large slave populations. In this case, our estimates could be picking up the presence of slave-descended populations rather than the malaria-related effects. To mitigate this concern, we augment our regression setup in two ways. First, we control

external validation of our measure. As in Sub-Saharan Africa, the sustained presence of Plasmodium falciparum implies that the climatic malaria potential represents (represented) a negative locational fundamental that shapes the spatial distribution of regional development.

Taken together, the results presented in this section provide strong evidence that our measure specifically captures the climatic potential for malaria transmission.[33] To further assuage concerns related to omitted variables biasing our results, we demonstrate that the estimates remain stable when we extend the set of controls to include all natural characteristics employed in Henderson et al. (2018)'s study which investigates determinants of the spatial distribution of economic activity worldwide (Tables C.1–C.2).[34] We additionally employ the procedure developed in Oster (2019) to more formally evaluate the robustness of our estimates to omitted variable bias. The results reported in Table C.3 show that for the grid-cell-level analysis, the influence of unobservables would need to be 1.5 to 5.1 times more important than all included observable variables to entirely attribute our results to the omission of unobserved factors. The relative influence of unobservables would have to be even greater to suppress the MSM effect in the individual-level analysis (Table C.4).

Additional robusteness tests

In additional robustness tests, we show that our results are not dependent on the time span used to construct our malaria suitability measure. Using monthly data over the period 1901—2010 leaves the estimates almost unchanged (Tables C.5–C.6). Coefficients, presented in Tables C.7–C.8, also remain similar if we use satellite-derived monthly temperature data between 2000–2010 from NASA EOSDIS (2015). This alleviates concerns related our results could be biased due to the spatial and temporal interpolation of the CRU climate data.[35] The use of a alternative standard error clustering approaches (Tables

for the municipality's share of total population that is either black or mulatto (our proxy for African descendance). Second, we directly control for the respondents' ancestry by including separate indicator variables for the race (as reported in the census and categorized into 5 groups). In both cases, the MSM point estimates remain virtually unchanged. Results are available upon request.

[33] A potential concern related to the analysis of Latin America is that temperatures in the high and low slave density regions are structurally different. In Appendix B.1 we show that this is not an issue.

[34] Specifically, we add controls for the length of growing period, land suitability for agriculture, ruggedness, a set of 14 biome indicators as well as separate dummy variables indicating whether a navigable river or lake lies within 25 kilometres of the grid cell centroids. In contrast to our setup, Henderson et al. (2018) further control for Kiszewski et al. (2004)'s malaria index. Our MSM point estimate remains stable irrespective of whether we include this index or not.

[35] The degree of spatial and temporal interpolation within Sub-Saharan Africa is likely to be high (relative to other regions) given the relative sparsity of weather stations (cf. Harris et al. (2014)).

C.9 and C.10) or the weighting procedures in the individual-level analysis (Table C.11) does not also change our findings. Furthermore, we show that our results are not driven by outliers. Applying the inverse hyperbolic sine transformation to our outcome variables or winsorizing them at the 5% level produces very similar results, both in terms of magnitude and statistical significance (Tables C.12–C.13).[36] This is also the case when we use alternative population data (Table C.14).

A further robustness-related worry is that the negative effect of the MSM is the result of cherry-picking functional forms or parameter values. Tables C.15–C.16 show that this is not the case. We obtain qualitatively equivalent results if we replace the adult mortality rate (as modelled in Mordecai et al. (2013)) with mortality models—some of which incorporate relative humidity in addition to temperature—from other exclusively laboratory-based studies.[37] We also obtain similar estimates if we use the (stand-alone) temperature-sensitive malaria suitability model proposed by Weiss et al. (2014).[38]

As a final specification check, we investigate whether our estimates are driven by a specific part of the MSM distribution. To this end, we employ the semiparametric regression approach developed by Robinson (1988). The results, represented graphically in Figure C.1, give no indication that any important nonlinear influence is neglected. The relationship between the outcome variables and the MSM are approximately linear and monotonically decreasing.

5 Malaria Suitability, Colonial Activity and Urbanization

In the last step of the analysis, we try to shed some light on the mechanisms underlying the malaria environment's influence on the spatial distribution of urbanization. Thereby, we focus on the locational choices of the European colonizers.

For Sub-Saharan Africa, the Scramble for Africa is seen as the critical juncture in the (modern) urbanization process (Stren and Halfani (2001), Freund (2007, p.65), Coquery-

[36] The inverse hyperbolic sine transformation approximates the natural logarithm and is—contrary to the logarithm—well defined around zero (Burbidge et al., 1988; Card and DellaVigna, Forthcoming). We obtain very similar results if we log-transform the variables after adding 1 or 0.01 to avoid losing observations with zeros (cf. Shenoy, 2018; Michalopoulos and Papaioannou, 2014). Results are available upon request.

[37] From Lunde et al. (2013b)—who compare different mortality models—we choose the two models that are (a) only based on laboratory studies, and (b) perform well in validation exercises.

[38] Compared to the Mordecai et al. (2013) model, the drawback of this suitability index is that not all aspects that influence R_0 are modelled as being temperature dependent (Weiss et al., 2014, p.6).

Vidrovitch (2005, p.4), Christopher and Tarver (1994), O'Connor (1983, p.32)). The colonizers established economic and administrative centers in which they built new infrastructure, such as power stations or new transport systems, and successively integrated the cities into the world market (Stren and Halfani (2001), Simon (1989), Coquery-Vidrovitch (2005, p. 328ff), Ady (1965, p. 11)). A crucial innovation was further the establishment of a system of market-based capitalism, characterized by wage labor and capital accumulation. This created a strong incentive for rural to urban migration (Rakodi (1997), Arthur (1991), Simon (1989)). As a consequence, many locations of intensive colonial economic activity have since evolved into major urban areas. Recent studies empirically validate these qualitative accounts and document that the location of European economic activity during the colonial period is an important determinant of today's spatial urbanization pattern (Huillery, 2009; Jedwab and Moradi, 2016).

Given the strong link between location of European colonial economic activity and location of cities today, a natural hypothesis is that the malaria environment influences the spatial urbanization pattern via a settlement-deterrent effect on the non-immune European colonizers (e.g., Webb (2014, p.25ff), Curtin (1985), Acemoglu et al. (2001)). To investigate the plausibility of this hypothesis, we first show that the local potential for malaria transmission influenced the location of European economic activity during the colonial era. We then assess, to what extent this effect can explain today's malaria-influenced spatial distribution of urbanization and how the effect varies over time. These last steps of our study are all carried out at the grid cell level.

Malaria Suitability and Location of European Colonial Centers

To investigate whether the local malaria transmission suitability deterred economic activity during the colonial era, we draw on the Oxford Regional Economic Atlas Africa (Ady, 1965).[39] This Atlas contains information on the locations of industries, electrical power production, mining activity, airports and train stations in the year 1956.[40] each of which we take as indicative of European economic activity (Ady, 1965, p. 11, 16).[41] We create a dummy variable that takes the value one if we find any of the formerly listed proxies for European activity in a grid cell, and zero otherwise. A cell is subsequently referred to as

[39]This data source is also used in Henderson et al. (2017) and Moradi (2005).

[40]For airports, information with respect to their location were gathered for the year 1953.

[41]It is important to note that the Atlas does not contain an exhaustive list of European settlements. Rather, it contains the subset of these locations that were characterized by substantial economic activity. These are also the centers which are more likely to have evolved into large urban areas.

‘European center’ whenever this dummy variable takes the value one. We further generate separate indicator variables for each of the individual sub-categories, i.e., the existence of industries, electrical power production, mining activity, airports or railway stations in the colonial era.

Table 8 column (1), documents that the local degree of malaria transmission suitability is statistically significantly negatively associated with the probability that a cell was home to a European center. Evaluated at the sample mean of 0.1, a one-standard deviation increase in the MSM reduces the likelihood of observing any colonial activity by 41 percent. We also find a negative effect when separately analyzing the relationship between the malaria potential and the likelihood that the Europeans built up industries or produced electrical power (columns (2)–(3)). On the other hand, there is no association between malaria suitability and the presence of mining centers. This indicates that the site of mining activity was determined by the (immobile) location of depletable minerals, which, in turn, is independent of local malaria transmission suitability. The fact that many of these mining centers developed into large urban areas implies that the constraint imposed by the malaria environment was not binding in these cases. The expected profit from resource extraction outweighed the malaria burden. Although of sizeable magnitude, the point estimate of

Table 8: Malaria Suitability and Colonial Presence

Dependent Variable:	Any Presence	Industries	Electrical Power	Mining	Railway Station	Airport
	(1)	(2)	(3)	(4)	(5)	(6)
MSM (SD)	-0.041***	-0.027**	-0.029***	-0.004	-0.012	-0.013**
	(0.015)	(0.011)	(0.009)	(0.007)	(0.010)	(0.006)
Climate controls	Yes	Yes	Yes	Yes	Yes	Yes
TseTse control	Yes	Yes	Yes	Yes	Yes	Yes
Geography controls	Yes	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	5,841	5,841	5,841	5,841	5,841	5,841
Mean LHS variable	0.101	0.038	0.040	0.028	0.038	0.031
Relative urban population[a]	12.710	17.503	13.790	6.253	11.328	12.284
R-squared	0.107	0.086	0.079	0.053	0.052	0.027

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at 2×2 degree grid cells reported in parentheses. All regressions control for the land surface area of the grid cells. Climate controls include temperature, precipitation and relative humidity, the squared terms of these variables and their first-order interactions as well as longitude, (absolute) latitude and a Tropics dummy. TseTse is the TseTse suitability index developed in Alsan (2015). Geography controls include distance to coast, elevation, waterway indicator, and caloric suitability.

[a] ‘Relative urban population’ reflects the urban population of the cells in with European activity relative to urban population in cells without activity.

MSM on the location of train stations—often lying on a line connecting resource extraction sites to ports—is statistically non-significant. On the other hand, airports in 1953, which we interpret as a more general proxy for the existence of European activity, are considerably less likely to be located in areas characterized by a high malaria transmission potential (column (6)).

Accounting for Colonial Economic Activity

Table 9 presents a mediation analysis. In column (1) we first re-run the grid-cell level regression of Table 2 Panel I, column (3), i.e., the specification in which we regress urbanization on the full set of controls. We then include the dummy variables that capture the existence of the different types of colonial centers into the set of control variables in column (2). The locational dummies absorb 38 percent of variation in the MSM.[42] In columns (3) and (4) we conduct the same exercise using night-time light intensity as outcome. Here, the MSM coefficient drops by 34 percent when we control for the presence of European activity. While these results cannot be interpreted as causal evidence, they suggest that the settlement pattern of the European colonizers (partly) explains why today's spatial distribution of urbanization is influenced by the local climatic malaria potential.[43] In the final step of this paper, we use panel data to provide further evidence consistent with this argument.

The Effect of the Malaria Environment on Urbanization over Time

Derived from census-based population counts, Jedwab and Moradi (2016) provide urban population estimates for the years 1890, 1900 and then from 1960 to 2010 at ten-year intervals.[44] We use this panel dataset—aggregated at the grid-cell$\times$census-year level—to investigate if our (time-invariant) malaria suitability measure exerts different effects on urban population over time. To this end, we interact the MSM with time-period fixed effects.

[42]The point estimates of the dummy variables that capture the existence of the different types of colonial centers are presented in Table D.1.

[43]To provide further evidence for the plausibility of this channel, we employ the presence of any Protestant or Catholic mission in 1924 (Roome, 1924) as an alternative proxy for European presence. The results in Table D.2 show that a high malaria transmission potential strongly deterred the establishment of missions. When additionally adding the presence of missions to the set of controls in Table 9, the size of the MSM coefficient is only marginally reduced. Running a horse race between the MSM and the presence of missions alone, the latter absorbs between 12 and 17 percent of the malaria effect.

[44]When censuses were conducted in different years, Jedwab and Moradi (2016) linearly interpolate or extrapolated the data across the years used in the analysis (see page A.5 in the Web Appendix of Jedwab and Moradi (2016)).

Table 9: Mediation: Malaria Suitability and Current-Day Grid-Cell-Level Urban Population

	Urban Population (SD)		Night-time Lights (SD)	
	(1)	(2)	(3)	(4)
MSM (SD)	-0.117**	-0.072*	-0.188***	-0.124***
	(0.051)	(0.040)	(0.061)	(0.044)
Country fixed effects	Yes	Yes	Yes	Yes
Climate controls	Yes	Yes	Yes	Yes
TseTse control	Yes	Yes	Yes	Yes
Geography controls	Yes	Yes	Yes	Yes
EU colonial activity controls	No	Yes	No	Yes
Observations	5,370	5,370	5,841	5,841
R-squared	0.054	0.251	0.173	0.326

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at 2×2 degree grid cells and reported in parentheses. Each coefficient reported in the table represents the point estimate for MSM (SD) obtained from running a separate version of regression Eq.(3). All regressions control for land surface area of the grid cells. Climate controls include temperature, precipitation and relative humidity, the squared terms of these variables and their first-order interactions as well as longitude, (absolute) latitude and a Tropics dummy. TseTse is the TseTse suitability index developed in Alsan (2015). Geography controls include distance to coast, elevation, waterway indicator, and caloric suitability.

In the subsequent analysis, we take 1890 as the base year. The time-period interacted coefficients of the MSM thus capture the differential effect of the malaria environment on urban population in a given period relative to the base year. The inclusion of grid cell fixed effects and time period dummies implies that time-invariant cell-specific differences as well as general time-specific changes are washed out. To account for the possibility that time-invariant characteristics may exhibit different effects in different time periods, we allow the full set of controls (i.e., climate, TseTse, and geography variables) to vary over time by interacting them with time-period fixed effects.[45] Figure 2 graphically depicts the point estimates along with the 90 percent confidence intervals. The figure illustrates that the effect of the climatic malaria environment has, relative to the base year 1890, increased over time, both in terms of size and statistical power. An immediate question is whether these increasing effects added to an existing baseline effect or whether there was no association between malaria suitability in the base year. We check this by running a

[45]Formally, the regression setup can be represented as:

$$y_{g,c,t} = \sum_{p=1900}^{2010} \psi_p \, MSM_{g,c} \times I_t^p + \sum_{p=1900}^{2010} \mathbf{C}'_{g,c} \, I_t^p \, \boldsymbol{\phi}_p + \gamma_g + \tau_t + \varepsilon_{c,i,t}. \tag{4}$$

The dependent variable $y_{g,c,t}$ is urban population of grid cell g, located in country c, in year t (standardized across the whole sample). Grid-cell-level fixed effects are represented by γ_g, time-period fixed effects by τ_t, and the idiosyncratic error term by $\varepsilon_{g,c,t}$. The coefficients ψ_p capture the additional effect of MSM in a given year relative to the base year 1890. The effects of the (time-interacted) climate and geography control variables are captured in vector $\boldsymbol{\phi}_p$. The standard errors are clustered at the grid-cell level.

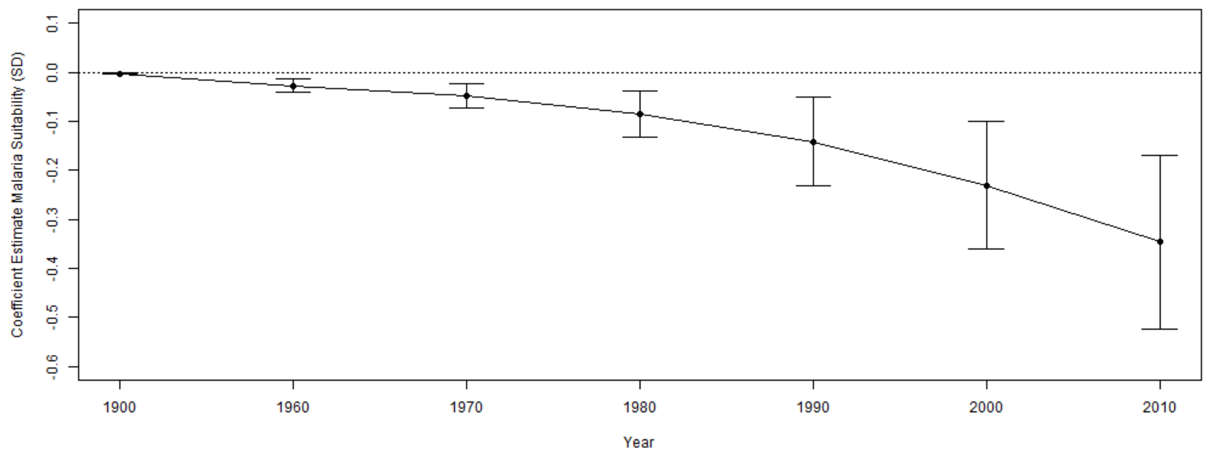

Figure 2: Figure depicts estimates of the time-period interacted effects of MSM relative to base year 1890 using the Jedwab and Moradi (2016).

cross-sectional regression akin to Eq.(3) using urban population data for the base year 1890 as the dependent variable. The point estimate is small and not statistically significant at conventional confidence levels (see Table D.3). This null result corroborates recent findings of Depetris-Chauvin and Weil (2018), who show that population densities and other measures of development were not influenced by the local malaria burden in the pre-colonial era. The finding that the malaria environment only started exerting a negative effect after the Scramble for Africa supports the view that the European colonizers played an important role in explaining why high malaria risk areas are less urbanized today. By influencing the location of colonial economic activity, the malaria environment also (partly) determined where modern urban areas emerged. As migration into these cities and internal population growth continues to rise, the effects of the malaria environment are compounded over time. It is important to note, however, that cities did exist in Africa prior to the arrival of the European colonizers. What our findings suggest is that the colonizer chose to establish economic activity in locations—either existing cities or newly established settlements—that were located in areas with a relatively low malaria transmission suitability.

6 Conclusion

In this study, we first derive an exogenous measure for the local climatic malaria transmission suitability. We then show that this locational fundamental influences the spatial distribution of urbanization and socioeconomic development in Sub-Saharan Africa. When interpreting these results it is important to keep in mind that they represent reduced-form effects, i.e., the net effect of all channels through which the malaria intensity measure affects outcomes. These may include direct health-related channels, such as increased mortality, potential reductions in labor productivity or lower human capital accumulation. Additionally, the climatic malaria potential can also influence today's local population densities and economic development through its effect on deeper fundamentals, such as institutions, culture or initial settlement choices. Crucially, malaria-induced disparities in these latter aspects—for the relevance of which we provide evidence—persist even if malaria is eradicated. Thus, while eradication of malaria will undoubtedly improve the living standards of people living in high transmission areas, (socio)-economic malaria-suitability-induced disparities across regions will likely persist.

References

Acemoglu, Daron, Simon Johnson, and James A. Robinson (2001) 'The colonial origins of comparative development: An empirical investigation.' *American Economic Review* 91(5), 1369–1401

Acemoglu, Daron, Simon Johnson, and James Robinson (2003) 'Disease and development in historical perspective.' *Journal of the European Economic Association* 1(2-3), 397–405

Acemoglu, Daron, Simon Johnson, and James Robinson (2005) 'The rise of Europe: Atlantic trade, institutional change, and economic growth.' *American Economic Review* 95(3), 546–579

Ady, Peter (1965) *Oxford Regional Economic Atlas Africa* (London: Oxford University Press)

Ali, Merima, Odd-Helge Fjeldstad, Boqian Jiang, and Abdulaziz B Shifa (2018) 'Colonial Legacy, State-building and the Salience of Ethnicity in Sub-Saharan Africa.' *The Economic Journal* 129(619), 1048–1081

Alsan, Marcella (2015) 'The effect of the TseTse fly on African development.' *American Economic Review* 105(1), 382–410

Amante, Christopher, and Barry W. Eakins (2009) *ETOPO1 1 arc-minute global relief model: procedures, data sources and analysis* (Boulder, Colorado: National Geophysical Data Center)

Amerasinghe, Felix P., Priyanie H. Amerasinghe, J. S. Malik Peiris, and Robert A. Wirtz

(1991) ‘Anopheline ecology and malaria infection during the irrigation development of an area of the Mahaweli Project, Sri Lanka.’ *The American Journal of Tropical Medicine and Hygiene* 45(2), 226–235

Arnfield, A. John (2003) ‘Two decades of urban climate research: a review of turbulence, exchanges of energy and water, and the urban heat island.’ *International Journal of Climatology* 23(1), 1–26

Arthur, John A. (1991) ‘International labor migration patterns in West Africa.’ *African Studies Review* 34(3), 65–87

Bhatt, S., D. J. Weiss, E. Cameron, D. Bisanzio, B. Mappin, U. Dalrymple, ..., and P. W. Gething (2015) ‘The effect of malaria control on plasmodium falciparum in africa between 2000 and 2015.’ *Nature* 526(7572), 207–211

Bleakley, Hoyt (2010) ‘Malaria eradication in the Americas: A retrospective analysis of childhood exposure.’ *American Economic Journal: Applied Economics* 2(2), 1–45

Bosker, Maarten, Steven Brakman, Harry Garretsen, and Marc Schramm (2007) ‘Looking for multiple equilibria when geography matters: German city growth and the WWII shock.’ *Journal of Urban Economics* 61(1), 152 – 169

Brinkhoff, Thomas (2015) ‘City population.’ www.citypopulation.de

Burbidge, John B., Lonnie Magee, and A. Leslie Robb (1988) ‘Alternative transformations to handle extreme values of the dependent variable.’ *Journal of the American Statistical Association* 83(401), 123–127

Card, David, and Stefano DellaVigna (Forthcoming) ‘What do editors maximize? evidence from four economics journals.’ *The Review of Economics and Statistics* pp. 1–53

Carter, Richard, and Kamini N. Mendis (2002) ‘Evolutionary and historical aspects of the burden of malaria.’ *Clinical Microbiology Reviews* 15(4), 564–594

Center for International Earth Science Information Network - CIESIN - Columbia University and International Food Policy Research Institute - IFPRI and The World Bank and Centro Internacional de Agricultura Tropical - CIAT (2011) ‘Global rural-urban mapping project, version 1 (grumpv1): Population count grid’

Christopher, A. J., and James D. Tarver (1994) ‘Urbanization during colonial days in Sub-Saharan Africa.’ In *Urbanization in Africa: A Handbook,* ed. James D. Tarver (Greenwood Press)

Coquery-Vidrovitch, Catherine (2005) *The History of African Cities South of the Sahara*

(Markus Wiener Publishers)

Curtin, Philip D. (1985) 'Medical knowledge and urban planning in tropical Africa.' *The American Historical Review* 90(3), 594–613

Cutler, David, Winnie Fung, Michael Kremer, Monica Singhal, and Tom Vogl (2010) 'Early-life malaria exposure and adult outcomes: Evidence from malaria eradication in India.' *American Economic Journal: Applied Economics* 2(2), 72–94

Davis, Donald R., and David E. Weinstein (2002) 'Bones, bombs, and break points: The geography of economic activity.' *American Economic Review* 92(5), 1269–1289

Dell, Melissa (2010) 'The Persistent Effects of Peru's Mining Mita.' *Econometrica* 78(6), 1863–1903

Dell, Melissa, Benjamin F. Jones, and Benjamin A. Olken (2012) 'Temperature shocks and economic growth: Evidence from the last half century.' *American Economic Journal: Macroeconomics* 4(3), 66–95

Depetris-Chauvin, Emilio, and David N. Weil (2018) 'Malaria and early African development: Evidence from the sickle cell trait.' *The Economic Journal* 128(610), 1207–1234

Easterly, William, and Ross Levine (2016) 'The European origins of economic development.' *Journal of Economic Growth* 21(3), 225–257

Ermert, Volker, Andreas H. Fink, Anne E. Jones, and Andrew P. Morse (2011) 'Development of a new version of the Liverpool Malaria Model. i. refining the parameter settings and mathematical formulation of basic processes based on a literature review.' *Malaria Journal* 10(1), 35

Flückiger, Matthias, and Markus Ludwig (2017) 'Malaria suitability, urbanization and persistence: Evidence from China over more than 2000 years.' *European Economic Review* 92, 146–160

Freund, Bill (2007) *The African City* (Cambridge University Press)

Gallup, John Luke, and Jeffrey D. Sachs (2001) 'The economic burden of malaria.' *The American Journal of Tropical Medicine and Hygiene* 64(1 suppl.), 85–96

Gallup, John Luke, Jeffrey D. Sachs, and Andrew D. Mellinger (1999) 'Geography and economic development.' *International Regional Science Review* 22(2), 179–232

Galor, Oded, and Ömer Özak (2015) 'Land Productivity and Economic Development: Caloric Suitability vs. Agricultural Suitability.' Technical Report, Brown University, Department of Economics

— (2016) 'The agricultural origins of time preference.' *American Economic Review* 106(10), 3064–3103

Gething, Peter W., Anand P. Patil, David L. Smith, Carlos A. Guerra, Iqbal R.F. Elyazar, Geoffrey L. Johnston, Andrew J. Tatem, and Simon I Hay (2011) 'A new world malaria map: Plasmodium falciparum endemicity in 2010.' *Malaria Journal* 10, 378–378

Guerra, C A, R W Snow, and S I Hay (2006) 'A global assessment of closed forests, deforestation and malaria risk.' *Annals of tropical medicine and parasitology* 100(3), 189–204

Hamlet, A, K Jean, W Perea, S Yactayo, J Biey, Kerkhove M Van, N Ferguson, and T Garske (2018) 'The seasonal influence of climate and environment on yellow fever transmission across africa.' *PLOS NEGLECTED TROPICAL DISEASES*

Harris, I., P.D. Jones, T.J. Osborn, and D.H. Lister (2014) 'Updated high-resolution grids of monthly climatic observations – the cru ts3.10 dataset.' *International Journal of Climatology* 34(3), 623–642

Hedrick, Philip (2011) 'Population genetics of malaria resistance in humans.' *Heredity* 107, 283–304

Henderson, J. Vernon, Adam Storeygard, and David N. Weil (2012) 'Measuring economic growth from outer space.' *American Economic Review* 102(2), 994–1028

Henderson, J. Vernon, Adam Storeygard, and Uwe Deichmann (2017) 'Has climate change driven urbanization in Africa?' *Journal of Development Economics* 124, 60 – 82

Henderson, J Vernon, Tim Squires, Adam Storeygard, and David Weil (2018) 'The global distribution of economic activity: Nature, history, and the role of trade1.' *The Quarterly Journal of Economics* 133(1), 357–406

Huillery, Elise (2009) 'History matters: The long-term impact of colonial public investments in French West Africa.' *American Economic Journal: Applied Economics* 1(2), 176–215

International Labour Organization (2015) 'Inventory of official national-level statistical definitions for rural/urban areas'

Jedwab, Remi, and Alexander Moradi (2016) 'The permanent effects of transportation revolutions in poor countries: Evidence from Africa.' *The Review of Economics and Statistics* 98(2), 268–284

Jedwab, Remi, Edward Kerby, and Alexander Moradi (2017) 'History, path dependence and development: Evidence from colonial railroads, settlers and cities in Kenya.' *Economic*

Journal 127(603), 1467–1494

Jeffery, Geoffrey M. (1976) 'Malaria control in the twentieth century.' *The American Journal of Tropical Medicine and Hygiene* 25(3), 361–371

Killeen, Gerry F., F.Ellis McKenzie, Brian D. Foy, Claus Bøgh, and John C. Beier (2001) 'The availability of potential hosts as a determinant of feeding behaviours and malaria transmission by African mosquito populations.' *Transactions of the Royal Society of Tropical Medicine and Hygiene* 95(5), 469 – 476

Kiszewski, A., A. Mellinger, A. Spielman, P. Malaney, S. E. Sachs, and J. Sachs (2004) 'A global index representing the stability of malaria transmission.' *American Journal of Tropical Medicine and Hygiene* 70(5), 486–498

Loy, Dorothy E., Weimin Liu, Yingying Li, Gerald H. Learn, Lindsey J. Plenderleith, Sesh A. Sundararaman, Paul M. Sharp, and Beatrice H. Hahn (2017) 'Out of africa: origins and evolution of the human malaria parasites plasmodium falciparum and plasmodium vivax.' *International Journal for Parasitology* 47(2), 87 – 97. Molecular Approaches to Malaria 2016 (MAM2016)

Lucas, Adrienne M. (2010) 'Malaria eradication and educational attainment: Evidence from paraguay and sri lanka.' *American Economic Journal: Applied Economics* 2(2), 46–71

Lucas, Adrienne M. (2013) 'The Impact of Malaria Eradication on Fertility.' *Economic Development and Cultural Change* 61(3), 607 – 631

Lunde, Torleif Markussen, Diriba Korecha, Eskindir Loha, Asgeir Sorteberg, and Bernt Lindtjørn (2013a) 'A dynamic model of some malaria-transmitting anopheline mosquitoes of the Afrotropical region. i. model description and sensitivity analysis.' *Malaria Journal* 12(1), 28

Lunde, Torleif Markussen, Mohamed Nabie Bayoh, and Bernt Lindtjørn (2013b) 'How malaria models relate temperature to malaria transmission.' *Parasites & Vectors* 6(1), 20

Martens, WJM, TH Jetten, J Rotmans, and LW Niessen (1995) 'Climate change and vector-borne diseases: A global modelling perspective.' *Global Environmental Change* 5(3), 195 – 209

McNeill, J. R. (2010) *Mosquito Empires: Ecology and War in the Greater Caribbean, 1620-1914 (New Approaches to the Americas)* (Cambridge University Press)

Michaels, Guy, and Ferdinand Rauch (2018) 'Resetting the Urban Network: 117-2012.' *The Economic Journal* 128(608), 378–412

Michalopoulos, Stelios, and Elias Papaioannou (2013) 'Pre-colonial ethnic institutions and contemporary african development.' *Econometrica* 81(1), 113–152

_ (2014) 'National institutions and subnational development in Africa.' *The Quarterly Journal of Economics* 129(1), 151–213

Miguel, Edward, and Gerard Roland (2011) 'The long-run impact of bombing vietnam.' *Journal of Development Economics* 96(1), 1 – 15

Moradi, Alexander (2005) 'Ernährung, wirtschaftliche Entwicklung und Bürgerkriege in Afrika südlich der Sahara (1950-2000).' PhD dissertation, University of Tübingen

Mordecai, Erin A, Jeremy Cohen, Michelle V Evans, Prithvi Gudapati, Leah R Johnson, Catherine A. Lippi, Kerri Miazgowicz, Courtney C Murdock, Jason R Rohr, Sadie J Ryan, Van Savage, Marta Shocket, Anna Stewart Ibarra, Matthew B Thomas, and Daniel P Weikel (2017) 'Detecting the impact of temperature on transmission of zika, dengue and chikungunya using mechanistic models.' *bioRxiv*

Mordecai, Erin A., Krijn P. Paaijmans, Leah R. Johnson, Christian Balzer, Tal Ben-Horin, Emily de Moor, Amy McNally, Samraat Pawar, Sadie J. Ryan, Thomas C. Smith, and Kevin D. Lafferty (2013) 'Optimal temperature for malaria transmission is dramatically lower than previously predicted.' *Ecology Letters* 16(1), 22–30

Motamed, Mesbah J., Raymond J.G.M. Florax, and William A. Masters (2014) 'Agriculture, transportation and the timing of urbanization: Global analysis at the grid cell level.' *Journal of Economic Growth* 19(3), 339–368

Murdock, George Peter (1959) *Africa: Its People and Their Culture History* (NewYork:: McGraw-Hill)

Mutuku, Francis M., Charles H. King, Peter Mungai, Charles Mbogo, Joseph Mwangangi, Eric M. Muchiri, Edward D. Walker, and Uriel Kitron (2011) 'Impact of insecticide-treated bed nets on malaria transmission indices on the south coast of kenya.' *Malaria Journal* 10(1), 356

Ngwa, Gideon A. (2006) 'On the population dynamics of the malaria vector.' *Bulletin of Mathematical Biology* 68(8), 2161–2189

Nunn, Nathan, and Diego Puga (2012) 'Ruggedness: The blessing of bad geography in Africa.' *The Review of Economics and Statistics* 94(1), 20–36

O'Connor, Anthony (1983) *The African City* (Hutchinson)

Oster, Emily (2019) 'Unobservable selection and coefficient stability: Theory and evidence.'

Journal of Business & Economic Statistics 37(2), 187–204

Piel, Frédéric B, Anand P Patil, Rosalind E Howes, Oscar A Nyangiri, Peter W Gething, Mewahyu Dewi, William H Temperley, Thomas N Williams, David J Weatherall, and Simon I Hay (2013) 'Global epidemiology of sickle haemoglobin in neonates: a contemporary geostatistical model-based map and population estimates.' *Lancet (London, England)* 381(9861), 142–151

Putterman, Louis, and David N. Weil (2010) 'Post-1500 population flows and the long-run determinants of economic growth and inequality.' *The Quarterly Journal of Economics* 125(4), 1627–1682

Rakodi, Carole (1997) 'Global forces, urban change, and urban management in Africa.' In *The Urban Challenge in Africa,* ed. Carole Rakodi (United Nations University Press)

Rappaport, Jordan, and Jeffrey D Sachs (2003) 'The United States as a Coastal Nation.' *Journal of Economic Growth* 8(1), 5–46

Redding, Stephen J., Daniel M. Sturm, and Nikolaus Wolf (2010) 'History and industry location: Evidence from German airports.' *Review of Economics and Statistics* 93(3), 814–831

Robinson, Peter M. (1988) 'Root-n-consistent semiparametric regression.' *Econometrica* 56(4), 931–954

Roome, William R.M (1924) 'Ethnographic survey of Africa: Showing the tribes and languages; also the stations of missionary societies [map].' 1:5,977,382

Rosenthal, Stuart S., and William C. Strange (2008) 'The attenuation of human capital spillovers.' *Journal of Urban Economics* 64(2), 373 – 389

Saiz, Albert (2010) 'The Geographic Determinants of Housing Supply*.' *The Quarterly Journal of Economics* 125(3), 1253–1296

Seyoum, A, F Balcha, M Balkew, A Ali, and T Gebre-Michael (2002) 'Impact of cattle keeping on human biting rate of anopheline mosquitoes and malaria transmission around Ziway, Ethiopia.' *East Afr Med J* 79(9), 485–490

Shenoy, Ajay (2018) 'Regional development through place-based policies: Evidence from a spatial discontinuity.' *Journal of Development Economics* 130, 173 – 189

Simon, David (1989) 'Colonial cities, postcolonial Africa and the world economy: a reinterpretation.' *International Journal of Urban and Regional Research* 13(1), 68–91

Smith, David L., and Nick Ruktanonchai (2010) 'Progress in modelling malaria transmission.' In *Modelling Parasite Transmission and Control,* ed. Edwin Michael and Robert C. Spear, vol. 673 of *Advances in Experimental Medicine and Biology* (Springer New York) pp. 1–12

Stren, Richard, and Mohamed Halfani (2001) 'The cities of Sub-Saharan Africa: From dependency to marginality.' In 'Handbook of Urban Studies' (SAGE Publications Ltd)

Tuno, Nobuko, Wilberforce Okeka, Noboru Minakawa, Masahiro Takagi, and Guiyun Yan (2005) 'Survivorship of anopheles gambiae sensu stricto (diptera: Culicidae) larvae in Western Kenya highland forest.' *Journal of Medical Entomology* 42(3), 270–277

Vittor, Amy Yomiko, Rober H. Gilman, James Tielsch, Gregory Glass, Tim Shields, Wagner Sanchez Lozano, Viviana Pinedo-Cancino, and Jonathan A. Patz (2006) 'The effect of deforestation on the human-biting rate of Anopheles Darlingi, the primary vector of falciparum malaria in the Peruvian amazon.' *The American Journal of Tropical Medicine and Hygiene* 74(1), 3–11

Wan, Zhengming, S. Hook, and G. Hulley (2015) 'Mod11c3 modis/terra land surface temperature/emissivity monthly l3 global 0.05deg cmg v006.' Technical Report Accessed 5 July 2020, NASA EOSDIS Land Processes DAAC

Webb, James L.A. (2008) *Humanity's Burden: A Global History of Malaria* (Cambridge University Press)

Webb, James L.A. (2014) *The Long Struggle Against Malaria in Tropical Africa* (Cambridge University Press)

Weiss, Daniel J., Samir Bhatt, Bonnie Mappin, Thomas P. Van Boeckel, David L. Smith, Simon I. Hay, and Peter W. Gething (2014) 'Air temperature suitability for Plasmodium falciparum malaria transmission in Africa 2000-2012: a high-resolution spatiotemporal prediction.' *Malaria Journal* 13(1), 171

Appendices

A Data and Descriptive Statistics

Table A.1: DHS variable construction

Variable Name	DHS variable(s)	Construction	Explanation
Secondary Education	v106	v106>1	
Tertiary Education	v106	v106=3	
Wealth Index	v190	Taken directly from DHS	Details on construction of index can be found here: `http://bit.do/DHSwealthindex`.
Modern Sector Occupation	v717	v717 $\notin$ 4,5	Following occupations are *not* classified as belonging to the modern (industry and service) sector: (4) Agricultural - self employed; (5) Agricultural - employee. Agricultural categories also include fishermen, foresters and hunters. For all individuals without employment, the value is missing
Urban Residency	v102	v102=1	

Table A.2: DHS data used to construct dataset employed in main part (Table 3)

Country	Phase	Year	Country	Phase	Year
Angola	5	2011	Malawi	6	2015-2016
Benin	6	2011-12	Mali	6	2012-2013
Burkina Faso	6	2010	Mozambique	6	2011
Burundi	6	2010	Namibia	6	2013
Congo Democratic Republic	6	2013-14	Nigeria	6	2013
Cote d'Ivoire	6	2011-12	Rwanda	7	2014
Cameroon	6	2011	Senegal	6	2010-2011
Ethiopia	6	2011	Sierra Leone	6	2013
Gabon	6	2012	Swaziland	5	2006-2007
Ghana	7	2014	Tanzania	6	2010
Guinea	6	2012	Togo	6	2013
Kenya	7	2014	Uganda	6	2011
Lesotho	7	2014	Zambia	6	2013-2014
Liberia	6	2013	Zimbabwe	6	2010-2011

Table A.3: Data used to construct dataset employed in descriptive part (Table 1). AIS, DHS or MIS with Malaria Rapid Diagnostic Test Results.

Country	Phase	Year	Country	Phase	Year
Angola	MIS6	2011	Malawi	MIS 7	2014
Benin	DHS6	2011-12	Mali	DHS 6	2012-2013
Burkina Faso	MIS 7	2014	Mozambique	DHS 6	2011
Burundi	DHS 6	2010	Nigeria	MIS 6	2010
Congo Democratic Republic	DHS 6	2013-14	Rwanda	DHS 7	2014
Cote d'Ivoire	DHS 6	2011-12	Senegal	DHS 6	2010-2011
Ghana	DHS 7	2014	Tanzania	AIS 6	2011-2012
Guinea	DHS 6	2012	Togo	DHS 6	2013
Kenya	MIS 7	2015	Uganda	DHS 6	2011
Liberia	MIS 6	2011			

Table A.4: Descriptive Statistics Grid-Cell Level: Key Variables

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**Raw mean**[a]	**Obs.**
	Urban Population and Nighttime light Intensity					
Jedwab and Moradi (2016)	0	1	-0.140	36.950	37,865	5,370
DMSP	0	1	-0.208	29.998	1,172	5,841
	Malaria Suitability Measure					
MSM (SD)	0	1	-2.660	1.245	221.189	5,841

Notes: [a] 'Raw mean' refers to the mean of the untransformed, i.e., non-standardized, outcome and malaria suitability index variables.

Table A.5: Descriptive Statistics Individual Level: Key Variables

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**Raw mean**[a]	**Obs.**
	Outcome Variables					
Secondary Education	0	1	-0.767	1.305	0.370	522,538
Tertiary Education	0	1	-.0237	4.219	0.053	522,538
HH Wealth Index	0	1	-1.387	1.416	2.980	287,858
Modern Sector Occupation	0	1	-1.059	0.944	0.529	340,901
HH Urban Residence	0	1	-0.732	1.367	0.349	522,538
	Malaria Suitability Measure					
MSM (SD)	0	1	-2.324	1.295	208.513	522,538

Notes: [a] 'Raw mean' refers to the mean of the untransformed, i.e., non-standardized, outcome and malaria suitability index variables.

B Background

Plasmodium Falciparum Malaria in the Americas

The Americas were first populated by humans via the Bering Strait land bridge, i.e., a relatively cold region, about 16,500 years ago (Goebel et al., 2008). The Plasmodium falciparum parasite, which had evolved in Africa over thousands of years and whose survival is critically dependent on a sufficiently high air temperature, did not reach the Americas in the pre-Columbian era due to the climatic barrier along the original population route (Webb (2014, p.66 ff), Carter and Mendis (2002)). Even though the Plasmodium falciparum parasite was absent, the climatic suitability for its transmission is very high in large parts of the continent. Furthermore, several Anopheles species capable of transmitting the parasite, such as the Anopheles darlingi, were already endemic in pre-Columbian America (Marinotti et al., 2013). Consequently, the central factor preventing the spread of malaria falciparum was the absence of the parasite itself.

This changed with the start of the slave trade in the 16th century when the Plasmodium falciparum parasite was introduced directly from Africa (Webb (2008, p. 73), Carter and Mendis (2002)). In contrast to Africa, however, the geographical diffusion of malaria falciparum was not pervasive. The limited spread was mainly due to the fact that the Anopheles vectors native to the Americas are not as efficient in transmitting malaria as the African vectors, particularly the Anopheles gambiae (McNeill (2010, p. 55) and Carter and Mendis (2002)). As a consequence, only regions that saw a large influx of infectious and immune humans, i.e., African slaves, experienced an intensive and stable transmission of malaria falciparum (Webb, 2008, p. 79).[46] A large concentration of African slaves was characteristic of the (sugar) plantation economy introduced by the European colonizers. The system relied heavily on slave labor, resulting in a continued inflow of African slaves into the sugar cane producing areas. This influx started in the mid-16th century and ended with the abolition of slavery around 1850 (Webb (2008, p. 73, p. 77), McNeill (2010, p. 23 ff), Curtin (1968)). Areas which were dominated by sugar plantations became highly malarious, which lead to a dramatic increase in mortality among European colonizers

[46] African slaves were not only less susceptible to malaria infection, but also much more likely to carry the Plasmodium falciparum parasite (e.g., Molina-Cruz and Barillas-Mury (2014)). This created a (locally confined) stable transmission environment despite the presence of relatively inefficient vectors. Stable transmission requires a reservoir of immune hosts, especially when vector efficiency is low. Otherwise, infected individuals die before the ineffective vector is able to infect further individuals.

(Webb, 2008, p. 73, p. 77).[47] This, in turn, led to an even greater dependence on African slaves (Curtin (1968), Webb (2008, p.77)).

Latin American regions that experienced a large influx of African slaves and, consequently, a high malaria falciparum burden were Brazil and the Guianas (Webb (2008, p. 73, p. 85 ff.), McNeill (2010, p. 23), Mann (2011, p. 111,p. 303, p. 366), Curtin (1968)). In the rest of mainland Latin America—referred to as mainland Spanish Americas—mining constituted the most important economic sector.[48] Compared to sugar economies, the mining sector relied predominantly on Amerindian labor (Borucki et al., 2015). As a consequence, the less virulent Plasmodium vivax was the dominant malaria parasite in mainland Spanish Americas. This type of malaria exerted a considerably lower burden on the health of the non-immune population (Webb, 2008, p. 81).

Today, the vast majority of the Americas is free of malaria falciparum. In areas where Plasmodium falciparum has not been eradicated, endemicity is lower than 5 percent (Gething et al., 2011). The low transmission intensity is, apart from the relatively inefficient Anopheles vectors, the result of large scale eradication campaigns that have been implemented since the 1950s (e.g., Bleakley (2010) or Jeffery (1976)).

Once the Plasmodium falciparum parasite had been introduced to the Americas, the intensity of transmission was, as in Africa, influenced by the local climatic suitability for parasite and vector development. However, as outlined above, only regions that experienced a large influx of African slaves saw a stable transmission of malaria. Historical accounts indicate that areas that fulfilled these prerequisites are Brazil and the Guianas. To provide further evidence, we draw on data from `www.slavevoyages.org` and compute the absolute number as well as the density of African slaves (expressed as slaves per square kilometer) for different regions of mainland Latin America. The results are depicted in Table B.1. As can be seen, the overwhelming majority of slaves shipped to mainland Latin America were destined for Brazil. In terms of density, Brazil is second only to Dutch Guiana. British Guinea had a slave density comparable to Brazil, as did French Guiana. On the other hand, the density in mainland Spanish Americas was substantially lower. In total, this

[47]Other characteristics associated with the sugar plantation economy—such as forest clearing—additionally facilitated the spread of malaria. The fact that the falciparum parasite out-competed the vivax parasite in areas where the Africans outnumbered the Europeans further increased the death toll amongst Europeans (Webb, 2008, p.77).

[48]Mainland Spanish Americas encompasses all of mainland Latin America, except Brazil and the Guineas. Compared to mainland Spanish America, the Spanish island colonies, Cuba and Puerto Rico, experienced large inflows of African slaves, due to their specialization in sugar production (see e.g., `www.slavevoyages.org`).

vast region received about 0.5 million African slaves.[49] The differences in slave density (slaves per square kilometer) depicted in Table B.1 are also reflected in population shares. Around the year 1800, black population accounted for approximately 4% of total population in mainland Spanish Americas (Sánchez-Albornoz, 1984), while it had surpassed 40% in Brazil (Marcílio, 1984). These patterns persist until today. For example, African descendants account for 10.6% of the total population in Colombia (which was part of mainland Spanish Americas, DANE (2005)). In Brazil, on the other hand, this share lay at 51.3% in 2010 (Censo Demográfico, 2010).

Table B.1: African Slaves Shipped to Latin America

Region	Slaves	Area (km^2)	Slave Density (Slaves per km^2)
(a) Dutch Guiana	294,653	163,821	1.798
(b) Brazil	4,864,374	8,514,215	0.571
(c) French Guiana	30,599	83,534	0.366
(d) British Guiana	72,685	214,970	0.338
(e) Mainland Spanish Americas	267,499	11,342,424	0.042

Note: Number of disembarked slaves retrieved from `www.slavevoyages.org`. Mainland Spanish Americas encompasses all of mainland Latin America except Brazil and the Guianas.

B.1 Temperature densities

A potential concern related to the analysis of Latin America is that temperatures in the high and low slave density regions are structurally different; particularity that the null results are due to a lack of variation in temperature in low-density countries. Figure B.1 shows that average temperatures are in lower in low-slave-density countries. However, the standard deviation is considerably higher in the former group of countries. To show that our results are not driven by the lower end of the temperature distribution, we restrict our analysis to grid cells in which the malaria transmission potential is greater than zero. The results remain qualitatively unchanged: the MSM exerts a negative effect in high-slave density areas while there is no influence in low-slave density regions.

[49] Borucki et al. (2015, p. 458)'s description illustrates the low density of African slaves in mainland Latin America: "[...] The dispersal of captives [slaves] over an immense geographic area, and the fact that their arrival occurred over a much longer time span than in any other major polity in the Americas, may have inhibited the emergence of both large and permanent regions of black demographic and cultural dominance during the three centuries of Spanish colonialism."

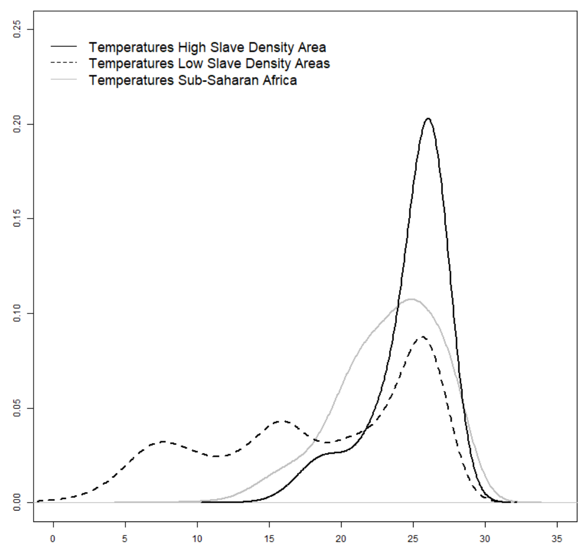

Figure B.1: Temperature densities across regions

C Robustness

Robustness: adding Henderson et al. (2018) controls

Table C.1: Adding Henderson et al. (2018) controls: Malaria Suitability and Current-Day Grid-Cell-Level Urban Population

	Current Day Urban Population (SD)			Night-time Lights (SD)		
	(1)	(2)	(3)	(4)	(5)	(6)
MSM (SD)	-0.135***	-0.127**	-0.124**	-0.148***	-0.173***	-0.162***
	(0.051)	(0.051)	(0.050)	(0.055)	(0.059)	(0.048)
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Climate controls	Yes	Yes	Yes	Yes	Yes	Yes
TseTse control	No	Yes	Yes	No	Yes	Yes
Geography controls	No	No	Yes	No	No	Yes
Observations	5,370	5,370	5,370	5,841	5,841	5,841
R-squared	0.051	0.051	0.082	0.155	0.156	0.200

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at 2×2 degree grid cells and reported in parentheses. Each coefficient reported in the table represents the point estimate for MSM (SD) obtained from running a separate version of regression Eq.(3). All regressions control for the land surface area of the grid cell. Climate controls include temperature, precipitation and relative humidity, the squared terms of these variables and their first-order interactions as well as longitude, (absolute) latitude and a Tropics dummy. TseTse is the TseTse suitability index developed in Alsan (2015). Geography controls include distance to coast, elevation, waterway indicator, and caloric suitability .

Table C.2: Adding Henderson et al. (2018) controls: Malaria Suitability and Individual-Level Socioeconomic Development

Dependent Variable:	Secondary Education (SD) (1)	Tertiary Education (SD) (2)	HH Wealth Index (SD) (3)	Non-Agricultural Occupation (SD) (4)	HH Urban Residence (SD) (5)
MSM (SD)	-0.059*	-0.098***	-0.175***	-0.157***	-0.205***
	(0.033)	(0.028)	(0.060)	(0.046)	(0.066)
Individual-level controls	Yes	Yes	Yes	Yes	Yes
Climate controls	Yes	Yes	Yes	Yes	Yes
TseTse control	Yes	Yes	Yes	Yes	Yes
Geography controls	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	522,538	522,538	287,858	340,901	287,858
R-squared	0.228	0.044	0.162	0.190	0.170

Note: Standard errors are clustered at the 0.5×0.5 degree grid cell level and are reported in parentheses. All regressions control for sex and age. Climate controls include temperature, precipitation and relative humidity as well as the squared terms of these variables and their first-order interactions. TseTse is the TseTse suitability index developed in Alsan (2015). Geography controls include distance to coast, elevation, waterway indicator, (absolute) latitude, longitude, a Tropics dummy and caloric suitability. Dependent variables are constructed from DHS data. See Section 3 and Table A.1 for more information.

Oster (2019) procedure to gauge influence of omitted variables

Table C.3: Oster (2019) Approach to Gauge Influence of Omitted Variables at Grid-Cell Level

Data Source of different urbanization measures	δ	R^2 baseline controls	R^2 full set of controls	R^2_{max}
Jedwab and Moradi (2016)	1.529	0.026	0.054	0.070
DMSP	3.609	0.097	0.173	0.225

Notes: The parameter δ represents how strong the influence of unobservables relative to observables would have to be in order to suppress the MSM effect. R^2 *baseline controls* captures the explanatory power obtained from the regressions that control for the set of climate variables. R^2 *full set of controls* reflects the explanatory power obtained from the regressions that control for the climate, TseTse, geography and country fixed effects. Following the recommendation of Oster (2019), we assume that the maximum achievable R-squared exceeds the R-squared obtained when including all observable covariates by 30%.

Table C.4: Oster (2019) Approach to Gauge Influence of Omitted Variables at Individual-Level

Dependent variable:	δ	R^2 baseline controls	R^2 full set of controls	R^2_{max}
Secondary Education (SD)	-2.458	0.185	0.224	0.292
Tertiary Education (SD)	4.286	0.039	0.042	0.055
Wealth Index (SD)	3.916	0.151	0.141	0.183
Modern Sector Occupation (SD)	38.371	0.182	0.178	0.232
Urban Residence (SD)	-3.381	0.154	0.141	0.184

Notes: The parameter δ represents how strong the influence of unobservables relative to observables would have to be in order to suppress the MSM effect. R^2 *baseline controls* captures the explanatory power obtained from the regressions that control for the set of climate variables. R^2 *full set of controls* reflects the explanatory power obtained from the regressions that control for the climate, TseTse, geography and country fixed effects. Following the recommendation of Oster (2019), we assume that the maximum achievable R-squared exceeds the R-squared obtained when including all observable covariates by 30%. A negative value of δ indicates that the inclusion of the full set of control variables increases the MSM point estimate compared to the regression that only accounts for the climate variables.

Robustness: time span climate variables (1901–2010)

Table C.5: Time Span Climate Variables Construction 1901–2010: Malaria Suitability and Current-Day Grid-Cell-Level Urban Population

	Current Day Urban Population (SD)			Night-time Lights (SD)		
	(1)	(2)	(3)	(4)	(5)	(6)
MSM (SD)	-0.133***	-0.126**	-0.120**	-0.137***	-0.163***	-0.187***
	(0.050)	(0.050)	(0.051)	(0.051)	(0.057)	(0.059)
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Climate controls	Yes	Yes	Yes	Yes	Yes	Yes
TseTse control	No	Yes	Yes	No	Yes	Yes
Geography controls	No	No	Yes	No	No	Yes
Observations	5,370	5,370	5,370	5,841	5,841	5,841
R-squared	0.051	0.052	0.054	0.152	0.154	0.171

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at 2×2 degree grid cells and reported in parentheses. Each coefficient reported in the table represents the point estimate for MSM (SD) obtained from running a separate version of regression Eq.(3). All regressions control for the land surface area of the grid cell. Climate controls include temperature, precipitation and relative humidity, the squared terms of these variables and their first-order interactions as well as longitude, (absolute) latitude and a Tropics dummy. TseTse is the TseTse suitability index developed in Alsan (2015). Geography controls include distance to coast, elevation, waterway indicator, and caloric suitability.

Table C.6: Time Span Climate Variables Construction 1901–2010: Malaria Suitability and Individual-Level Socioeconomic Development

Dependent Variable:	Secondary Education (SD)	Tertiary Education (SD)	HH Wealth Index (SD)	Non-Agricultural Occupation (SD)	HH Urban Residence (SD)
	(1)	(2)	(3)	(4)	(5)
MSM (SD)	-0.131***	-0.148***	-0.341***	-0.247***	-0.408***
	(0.039)	(0.041)	(0.079)	(0.053)	(0.092)
Individual-level controls	Yes	Yes	Yes	Yes	Yes
Climate controls	Yes	Yes	Yes	Yes	Yes
TseTse control	Yes	Yes	Yes	Yes	Yes
Geography controls	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes
Obs.	522,538	522,538	287,858	340,901	287,858
R-squared	0.225	0.042	0.140	0.178	0.142

Note: Standard errors are clustered at the DHS cluster level and are reported in parentheses. All regressions control for sex and age. Climate controls include temperature, precipitation and relative humidity as well as the squared terms of these variables and their first-order interactions. TseTse is the TseTse suitability index developed in Alsan (2015). Geography controls include distance to coast, elevation, waterway indicator, (absolute) latitude, longitude, a Tropics dummy and caloric suitability. Dependent variables are constructed from DHS data. See Section 3 and Table A.1 for more information.

Robustness: satellite-derived temperature data

Table C.7: Satellite-derived temperature data: Malaria Suitability and Current-Day Grid-Cell-Level Urban Population

	Current Day Urban Population (SD)			Night-time Lights (SD)		
	(1)	(2)	(3)	(4)	(5)	(6)
MSM (SD)	-0.125***	-0.128***	-0.144***	-0.155**	-0.174***	-0.198***
	(0.034)	(0.033)	(0.037)	(0.061)	(0.061)	(0.074)
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Climate controls	Yes	Yes	Yes	Yes	Yes	Yes
TseTse control	No	Yes	Yes	No	Yes	Yes
Geography controls	No	No	Yes	No	No	Yes
Observations	5,370	5,370	5,370	5,841	5,841	5,841
R-squared	0.048	0.048	0.053	0.150	0.155	0.169

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at 2×2 degree grid cells and reported in parentheses. Each coefficient reported in the table represents the point estimate for MSM (SD) obtained from running a separate version of regression Eq.(3). All regressions control for the land surface area of the grid cell. Climate controls include temperature, precipitation and relative humidity, the squared terms of these variables and their first-order interactions as well as longitude, (absolute) latitude and a Tropics dummy. TseTse is the TseTse suitability index developed in Alsan (2015). Geography controls include distance to coast, elevation, waterway indicator, and caloric suitability.

Table C.8: Satellite-derived temperature data: Malaria Suitability and Individual-Level Socioeconomic Development

Dependent Variable:	Secondary Education (SD) (1)	Tertiary Education (SD) (2)	HH Wealth Index (SD) (3)	Non-Agricultural Occupation (SD) (4)	HH Urban Residence (SD) (5)
MSM (SD)	-0.102***	-0.106***	-0.172***	-0.120**	-0.231***
	(0.022)	(0.021)	(0.064)	(0.048)	(0.074)
Individual-level controls	Yes	Yes	Yes	Yes	Yes
Climate controls	Yes	Yes	Yes	Yes	Yes
TseTse control	Yes	Yes	Yes	Yes	Yes
Geography controls	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes
Obs.	522,538	522,538	287,858	340,901	287,858
R-squared	0.226	0.043	0.142	0.177	0.143

Note: Standard errors are clustered at the DHS cluster level and are reported in parentheses. All regressions control for sex and age. Climate controls include temperature, precipitation and relative humidity as well as the squared terms of these variables and their first-order interactions. TseTse is the TseTse suitability index developed in Alsan (2015). Geography controls include distance to coast, elevation, waterway indicator, (absolute) latitude, longitude, a Tropics dummy and caloric suitability. Dependent variables are constructed from DHS data. See Section 3 and Table A.1 for more information.

Robustness: alternative clustering approaches

Table C.9: Conley (1999) Standard Errors: Malaria Suitability and Current-Day Grid-Cell-Level Urban Population

	Current Day Urban Population (SD)			Night-time Lights (SD)		
	(1)	(2)	(3)	(4)	(5)	(6)
MSM (SD)	-0.135***	-0.127**	-0.117**	-0.148***	-0.173***	-0.188***
	[0.051]	[0.051]	[0.052]	[0.051]	[0.055]	[0.057]
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Climate controls	Yes	Yes	Yes	Yes	Yes	Yes
TseTse control	No	Yes	Yes	No	Yes	Yes
Geography controls	No	No	Yes	No	No	Yes
Observations	5,370	5,370	5,370	5,841	5,841	5,841
R-squared	0.051	0.051	0.054	0.155	0.156	0.173

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors computed using the approach of Conley (1999) reported in brackets (cut-off 2 degrees). Each coefficient reported in the table represents the point estimate for MSM (SD) obtained from running a separate version of regression Eq.(3). All regressions control for the land surface area of the grid cell. Climate controls include temperature, precipitation and relative humidity, the squared terms of these variables and their first-order interactions as well as longitude, (absolute) latitude and a Tropics dummy. TseTse is the TseTse suitability index developed in Alsan (2015). Geography controls include distance to coast, elevation, waterway indicator, and caloric suitability.

Table C.10: Alternative Standard Errors: Malaria Suitability and Individual-Level Socioeconomic Development

Dependent Variable:	Secondary Education (SD) (1)	Tertiary Education (SD) (2)	HH Wealth Index (SD) (3)	Non-Agricultural Occupation (SD) (4)	HH Urban Residence (SD) (5)
MSM (SD)	-0.121***	-0.145***	-0.332***	-0.243***	-0.399***
	(0.016)	(0.018)	(0.027)	(0.024)	(0.034)
Individual-level controls	Yes	Yes	Yes	Yes	Yes
Climate controls	Yes	Yes	Yes	Yes	Yes
TseTse control	Yes	Yes	Yes	Yes	Yes
Geography controls	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	522,538	522,538	287,858	340,901	287,858
R-squared	0.224	0.042	0.140	0.178	0.141

Note: Standard errors are clustered at DHS cluster level and are reported in parentheses. All regressions control for sex and age. Climate controls include temperature, precipitation and relative humidity as well as the squared terms of these variables and their first-order interactions. TseTse is the TseTse suitability index developed in Alsan (2015). Geography controls include distance to coast, elevation, waterway indicator, (absolute) latitude, longitude, a Tropics dummy and caloric suitability. Dependent variables are constructed from DHS data. See Section 3 and Table A.1 for more information.

Robustness: alternative weighting DHS data

Table C.11: Varying weights: Malaria Suitability and Individual-Level Socioeconomic Development

Dependent Variable:	Secondary Education (SD) (1)	Tertiary Education (SD) (2)	HH Wealth Index (SD) (3)	Non-Agricultural Occupation (SD) (4)	HH Urban Residence (SD) (5)
	Panel A: No weights				
MSM (SD)	-0.086***	-0.103***	-0.164***	-0.211***	-0.259***
	(0.031)	(0.030)	(0.043)	(0.062)	(0.062)
	Panel B: Equal weighting				
MSM (SD)	-0.148***	-0.154***	-0.291***	-0.405***	-0.430***
	(0.034)	(0.033)	(0.057)	(0.075)	(0.082)
	Panel B: Population-size weighted				
MSM (SD)	-0.080*	-0.110**	-0.283***	-0.338***	-0.395***
	(0.046)	(0.047)	(0.065)	(0.091)	(0.107)
Individual-level controls	Yes	Yes	Yes	Yes	Yes
Climate controls	Yes	Yes	Yes	Yes	Yes
TseTse control	Yes	Yes	Yes	Yes	Yes
Geography controls	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	522,538	522,538	287,858	340,901	287,858

Note: Standard errors are clustered at the 0.5×0.5 degree grid cell level and are reported in parentheses. All regressions control for sex and age. Climate controls include temperature, precipitation and relative humidity as well as the squared terms of these variables and their first-order interactions. TseTse is the TseTse suitability index developed in Alsan (2015). Geography controls include distance to coast, elevation, waterway indicator, (absolute) latitude, longitude, a Tropics dummy and caloric suitability. Dependent variables are constructed from DHS data.

Robustness: reducing influence of outliers

Table C.12: Inverse hyperbolic sine transformation: Malaria Suitability and Current-Day Grid-Cell-Level Urban Population

	Log Day Urban Population			Log Night-time Lights		
	(1)	(2)	(3)	(4)	(5)	(6)
MSM (SD)	-0.873***	-0.795***	-0.666***	-0.824***	-0.949***	-0.801***
	(0.222)	(0.236)	(0.238)	(0.165)	(0.186)	(0.186)
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Climate controls	Yes	Yes	Yes	Yes	Yes	Yes
TseTse control	No	Yes	Yes	No	Yes	Yes
Geography controls	No	No	Yes	No	No	Yes
Observations	5,370	5,370	5,370	5,841	5,841	5,841
R-squared	0.229	0.230	0.244	0.376	0.378	0.396

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at 2×2 degree grid cells and reported in parentheses. Each coefficient reported in the table represents the point estimate for MSM (SD) obtained from running a separate version of regression Eq.(3). All regressions control for the land surface area of the grid cell. Climate controls include temperature, precipitation and relative humidity, the squared terms of these variables and their first-order interactions as well as longitude, (absolute) latitude and a Tropics dummy. TseTse is the TseTse suitability index developed in Alsan (2015). Geography controls include distance to coast, elevation, waterway indicator, and caloric suitability. LHS variables are transformed using the inverse hyperbolic sine function.

Table C.13: Winsorizing LHS at the 5% level: Malaria Suitability and Current-Day Grid-Cell-Level Urban Population

	Current Day Urban Population (SD)			Night-time Lights (SD)		
	(1)	(2)	(3)	(4)	(5)	(6)
MSM (SD)	-0.159***	-0.140***	-0.117**	-0.200***	-0.245***	-0.246***
	(0.045)	(0.047)	(0.048)	(0.059)	(0.064)	(0.061)
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Climate controls	Yes	Yes	Yes	Yes	Yes	Yes
TseTse control	No	Yes	Yes	No	Yes	Yes
Geography controls	No	No	Yes	No	No	Yes
Observations	5,370	5,370	5,370	5,841	5,841	5,841
R-squared	0.213	0.214	0.224	0.349	0.353	0.376

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at 2×2 degree grid cells and reported in parentheses. Each coefficient reported in the table represents the point estimate for MSM (SD) obtained from running a separate version of regression Eq.(3). All regressions control for the land surface area of the grid cell. Climate controls include temperature, precipitation and relative humidity, the squared terms of these variables and their first-order interactions as well as longitude, (absolute) latitude and a Tropics dummy. TseTse is the TseTse suitability index developed in Alsan (2015). Geography controls include distance to coast, elevation, waterway indicator, and caloric suitability.

Robustness: alternative population data

Table C.14: Alternative Population Data: Malaria Suitability and Current-Day Grid-Cell-Level Urban Population

	Current Day Urban Population (SD) citypopulation.de			Current Day Urban Population (SD) CIESIN (2005)		
	(1)	(2)	(3)	(4)	(5)	(6)
MSM (SD)	-0.135***	-0.127**	-0.117**	-0.148***	-0.173***	-0.188***
	(0.051)	(0.051)	(0.051)	(0.055)	(0.059)	(0.061)
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Climate controls	Yes	Yes	Yes	Yes	Yes	Yes
TseTse control	No	Yes	Yes	No	Yes	Yes
Geography controls	No	No	Yes	No	No	Yes
Observations	5,370	5,370	5,370	5,841	5,841	5,841
R-squared	0.051	0.051	0.054	0.155	0.156	0.173

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at 2×2 degree grid cells and reported in parentheses. Each coefficient reported in the table represents the point estimate for MSM (SD) obtained from running a separate version of regression Eq.(3). All regressions control for the land surface area of the grid cell. Climate controls include temperature, precipitation and relative humidity, the squared terms of these variables and their first-order interactions as well as longitude, (absolute) latitude and a Tropics dummy. TseTse is the TseTse suitability index developed in Alsan (2015). Geography controls include distance to coast, elevation, waterway indicator, and caloric suitability. Brinkhoff (2015) provides census-based city-level population data. With the exception of Somalia, data are provided for all Sub-Saharan African countries. The years for which population data is available vary across countries. We use the most recent census year available, which lies between 1970 and 2014. CIESIN (2005) provides gridded estimates of the total population for the year 2000 at a spatial resolution of 1×1km. To obtain estimates for urban population, we overlay the population grid with an urban extent grid which is also provided by (CIESIN). Only population counts of the 1×1 km grid cells that lie within the urban extents are taken into account in the aggregation to the 0.5×0.5 degree grid-cell level.

Robustness: alternative malaria suitability measures

Table C.15: Robustness MSM: Malaria Suitability and Current-Day Grid-Cell-Level Urban Population

	Current Day Urban Population (SD)			Night-time Lights (SD)		
	(1)	(2)	(3)	(4)	(5)	(6)
MSM (SD)	-0.121**	-0.096*	-0.103*	-0.186***	-0.134**	-0.108**
	(0.060)	(0.051)	(0.056)	(0.070)	(0.055)	(0.044)
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Climate controls	Yes	Yes	Yes	Yes	Yes	Yes
TseTse control	Yes	Yes	Yes	Yes	Yes	Yes
Geography controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,370	5,370	5,370	5,841	5,841	5,841
R-squared	0.053	0.053	0.054	0.172	0.172	0.172
Modification MSM	Martens3	Bayoh-Parham	Weiss	Martens3	Bayoh-Parham	Weiss

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at 2×2 degree grid cells and reported in parentheses. Each coefficient reported in the table represents the point estimate for MSM (SD) obtained from running a separate version of regression Eq.(3). All regressions control for the land surface area of the grid cell. Climate controls include temperature, precipitation and relative humidity, the squared terms of these variables and their first-order interactions as well as longitude, (absolute) latitude and a Tropics dummy. TseTse is the TseTse suitability index developed in Alsan (2015). Geography controls include distance to coast, elevation, waterway indicator, and caloric suitability. In columns (1) and (4), the original vector mortality model is replaced by the 'Martens 3' model, in columns (2) and (5) by the 'Bayoh-Parham' model (Lunde et al., 2013). In columns (3) and (6), MSM is replaced by the temperature suitability model developed in Weiss et al. (2014).

Table C.16: Robustness MSM: Malaria Suitability and Individual-Level Socioeconomic Development

Dependent Variable:	Secondary Education (SD) (1)	Tertiary Education (SD) (2)	HH Wealth Index (SD) (3)	Non-Agricultural Occupation (SD) (4)	HH Urban Residence (SD) (5)
	Panel A: Martens 3				
MSM (SD)	-0.131***	-0.158***	-0.352***	-0.248***	-0.432***
	(0.044)	(0.044)	(0.089)	(0.060)	(0.102)
	Panel B: Bayoh-Parham				
MSM (SD)	-0.149***	-0.182***	-0.435***	-0.378***	-0.550***
	(0.050)	(0.049)	(0.101)	(0.069)	(0.114)
	Panel C: Weiss et al. (2014)				
MSM (SD)	-0.042***	-0.046***	-0.145***	-0.138***	-0.187***
	(0.016)	(0.014)	(0.035)	(0.028)	(0.040)
Climate controls	Yes	Yes	Yes	Yes	Yes
TseTse control	Yes	Yes	Yes	Yes	Yes
Geography controls	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes
Obs.	522,538	522,538	287,858	340,901	287,858

Note: Standard errors are clustered at the 0.5×0.5 degree grid cell level and are reported in parentheses. All regressions control for sex and age. Climate controls include temperature, precipitation and relative humidity as well as the squared terms of these variables and their first-order interactions. TseTse is the TseTse suitability index developed in Alsan (2015). Geography controls include distance to coast, elevation, waterway indicator, (absolute) latitude, longitude, a Tropics dummy and caloric suitability. Dependent variables are constructed from DHS data. See Section 3 and Table A.1 for more information.

Semiparametric regression approachRobinson (1988)

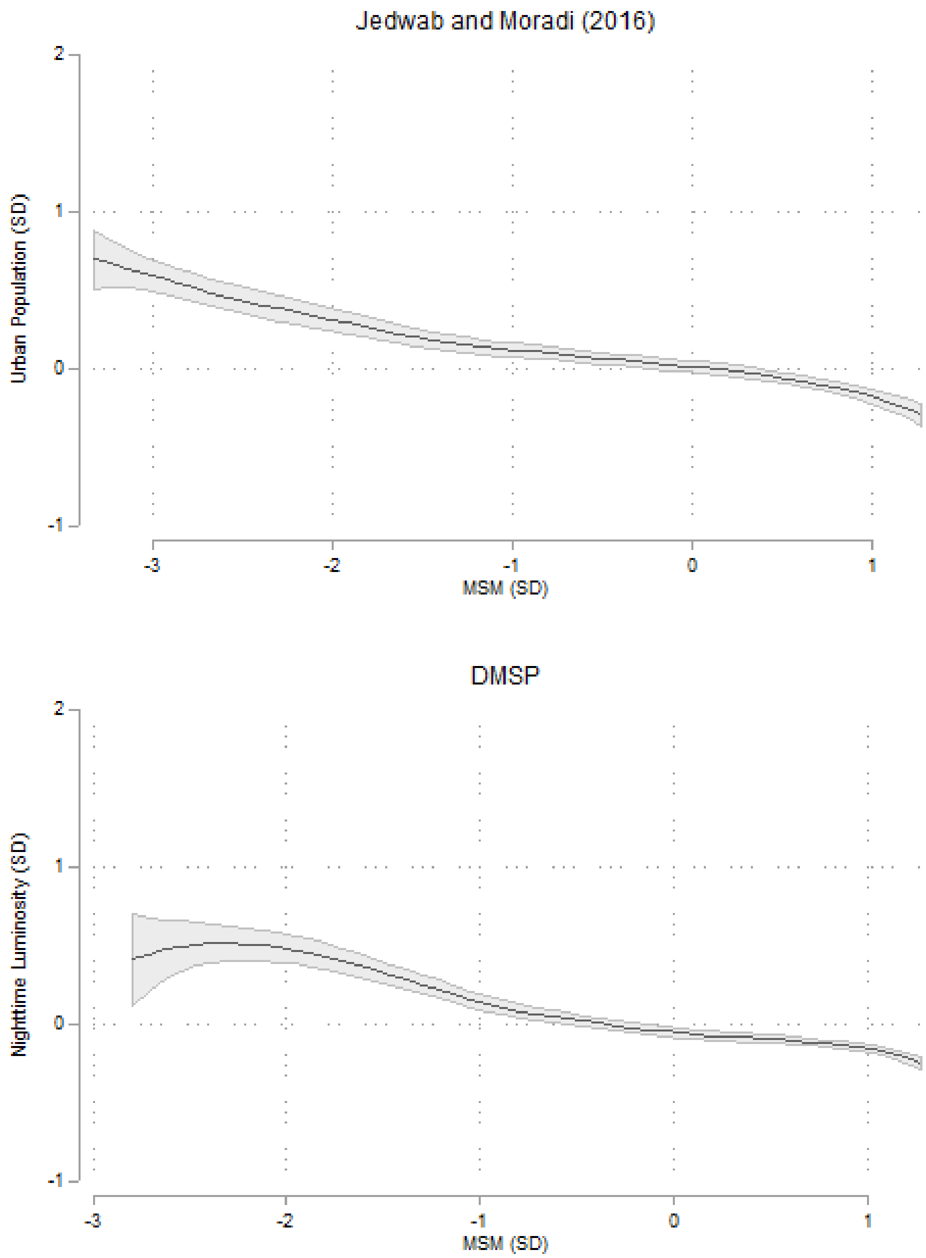

Figure C.1: Estimates are produced employing the semiparametric regression approach of Robinson (1988) with a second degree Epanechnikov kernel. The parametric part of the regression model includes the full set of climate, TseTse, Geography controls as well as country fixed effects (see Section 3). The shaded areas represent the 95 percent confidence bands.

D Supporting Evidence

Table D.1: Malaria Suitability: Controlling for Colonial Economic Activity (Covariates)

	Current Day Urban Population (SD)	Nighttime Lights (SD)
	(1)	(2)
MSM (SD)	-0.072*	-0.124***
	(0.040)	(0.044)
Industries	1.462***	1.480***
	(0.285)	(0.263)
Electrical Power	0.729***	0.487***
	(0.241)	(0.138)
Mining	-0.148	0.291
	(0.278)	(0.258)
Railway Station	0.771***	0.428***
	(0.262)	(0.124)
Airport	0.613***	0.270**
	(0.228)	(0.121)
Country fixed effects	Yes	Yes
Climate controls	Yes	Yes
TseTse control	Yes	Yes
Geography controls	Yes	Yes
Observations	5,370	5,370
Obs.	5,370	5,841
R-squared	0.251	0.326

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at 2×2 degree grid cells and reported in parentheses. Each coefficient reported in the table represents the point estimate for MSM (SD) obtained from running a separate version of regression Eq.(3). All regressions control for the landmass encompassed by the individual grid cells. Climate controls include temperature, precipitation and relative humidity, the squared terms of these variables and their first-order interactions as well as longitude, (absolute) latitude and a Tropics dummy. TseTse is the TseTse suitability index developed in Alsan (2015). Geography controls include distance to coast, elevation, waterway indicator, and caloric suitability.

Table D.2: Malaria Suitability: Controlling for Presence of Missions

	Mission	Current-Day Urban Population (SD)	Night-time Lights (SD)
	(1)	(2)	(3)
MSM	-0.032**	-0.066*	-0.119***
	(0.016)	(0.039)	(0.044)
Climate controls	Yes	Yes	Yes
TseTse control	Yes	Yes	Yes
Geography controls	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes
Missions and EU colonial activity control	No	Yes	No
R-squared	0.203	0.254	0.329

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at 2×2 degree grid cells and reported in parentheses. Each coefficient reported in the table represents the point estimate for MSM (SD) obtained from running a separate version of regression Eq.(3). All regressions control for the landmass encompassed by the individual grid cells. Climate controls include temperature, precipitation and relative humidity, the squared terms of these variables and their first-order interactions as well as longitude, (absolute) latitude and a Tropics dummy. TseTse is the TseTse suitability index developed in Alsan (2015). Geography controls include distance to coast, elevation, waterway indicator, and caloric suitability. 'Mission' represents an indicator variable that takes the value of one if a Protestant or Catholic mission is located in a given grid cell in the year 1924. Data is taken from Roome (1924)

Table D.3: Malaria Suitability and Pre-1900 Urbanization

	Urban Population 1890 (SD)
	(1)
MSM (SD)	-0.043
	(0.050)
Climate controls	Yes
TseTse control	Yes
Geography controls	Yes
Country fixed effects	Yes
Obs	5,370
R-squared	0.019

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at 2×2 degree grid cells and reported in parentheses. Each coefficient reported in the table represents the point estimate for MSM (SD) obtained from running a separate version of regression Eq.(3). All regressions control for land surface area of the grid cells. Climate controls include temperature, precipitation and relative humidity, the squared terms of these variables and their first-order interactions as well as longitude, (absolute) latitude and a Tropics dummy. TseTse is the TseTse suitability index developed in Alsan (2015). Geography controls include distance to coast, elevation, waterway indicator, and caloric suitability.

References

Alsan, Marcella (2015) 'The effect of the TseTse fly on African development.' *American Economic Review* 105(1), 382–410

Bleakley, Hoyt (2010) 'Malaria eradication in the Americas: A retrospective analysis of childhood exposure.' *American Economic Journal: Applied Economics* 2(2), 1–45

Borucki, Alex, David Eltis, and David Wheat (2015) 'Atlantic history and the slave trade to Spanish America.' *The American Historical Review* 120(2), 433

Brinkhoff, Thomas (2015) 'City population.' www.citypopulation.de

Carter, Richard, and Kamini N. Mendis (2002) 'Evolutionary and historical aspects of the burden of malaria.' *Clinical Microbiology Reviews* 15(4), 564–594

Center for International Earth Science Information Network - CIESIN - Columbia University and International Food Policy Research Institute - IFPRI and The World Bank and Centro Internacional de Agricultura Tropical - CIAT (2011a) 'Global rural-urban mapping project, version 1 (grumpv1): Population count grid'

_ (2011b) 'Global rural-urban mapping project, version 1 (grumpv1): Urban extents grid'

Conley, T. G. (1999) 'GMM estimation with cross sectional dependence.' *Journal of Econometrics* 92(1), 1–45

Curtin, Philip D. (1968) 'Epidemiology and the slave trade.' *Political Science Quarterly* 83(2), pp. 190–216

DANE (2005) 'La visibilización estadística de los grupos étnicos colombianos.' Technical Report, Departamento Administrativo Nacional de Estadística(DANE), República de Colombia

Gething, Peter W., Anand P. Patil, David L. Smith, Carlos A. Guerra, Iqbal R.F. Elyazar, Geoffrey L. Johnston, Andrew J. Tatem, and Simon I Hay (2011) 'A new world malaria map: Plasmodium falciparum endemicity in 2010.' *Malaria Journal* 10, 378–378

Goebel, Ted, Michael R. Waters, and Dennis H. O'Rourke (2008) 'The late pleistocene dispersal of modern humans in the Americas.' *Science* 319(5869), 1497–1502

Henderson, J Vernon, Tim Squires, Adam Storeygard, and David Weil (2018) 'The global distribution of economic activity: Nature, history, and the role of trade1.' *The Quarterly Journal of Economics* 133(1), 357–406

Jedwab, Remi, and Alexander Moradi (2016) 'The permanent effects of transportation revolutions in poor countries: Evidence from Africa.' *The Review of Economics and*

Statistics 98(2), 268–284

Jeffery, Geoffrey M. (1976) 'Malaria control in the twentieth century.' *The American Journal of Tropical Medicine and Hygiene* 25(3), 361–371

Lunde, Torleif Markussen, Mohamed Nabie Bayoh, and Bernt Lindtjørn (2013) 'How malaria models relate temperature to malaria transmission.' *Parasites & Vectors* 6(1), 20

Mann, Charles C. (2011) *1493: Uncovering the New World Columbus Created* (Knopf)

Marcílio, Maria Luiza (1984) 'The population of colonial Brazil.' In *The Cambridge History of Latin America:,* ed. Leslie Bethell, vol. 2 (Cambridge: Cambridge University Press) pp. 37–64

Marinotti, Osvaldo, Gustavo C. Cerqueira, Luiz Gonzaga Paula de Almeida, Maria Ines Tiraboschi Ferro, and Et. al. (2013) 'The genome of anopheles darlingi, the main neotropical malaria vector.' *Nucleic Acids Research* 15(41), 7387–7400

McNeill, J. R. (2010) *Mosquito Empires: Ecology and War in the Greater Caribbean, 1620-1914 (New Approaches to the Americas)* (Cambridge University Press)

Molina-Cruz, Alvaro, and Carolina Barillas-Mury (2014) 'The remarkable journey of adaptation of the Plasmodium falciparum malaria parasite to New World anopheline mosquitoes.' *Memórias do Instituto Oswaldo Cruz* 109(5), 662–667

Oster, Emily (2019) 'Unobservable selection and coefficient stability: Theory and evidence.' *Journal of Business & Economic Statistics* 37(2), 187–204

Robinson, Peter M. (1988) 'Root-n-consistent semiparametric regression.' *Econometrica* 56(4), 931–954

Roome, William R.M (1924) 'Ethnographic survey of Africa: Showing the tribes and languages; also the stations of missionary societies [map].' 1:5,977,382

Sánchez-Albornoz, Nicolás (1984) 'The population of colonial Spanish America.' In *The Cambridge History of Latin America:,* ed. Leslie Bethell, vol. 2 (Cambridge: Cambridge University Press) pp. 1–36

Webb, James L.A. (2008) *Humanity's Burden: A Global History of Malaria* (Cambridge University Press)

Webb, James L.A. (2014) *The Long Struggle Against Malaria in Tropical Africa* (Cambridge University Press)

Weiss, Daniel J., Samir Bhatt, Bonnie Mappin, Thomas P. Van Boeckel, David L. Smith, Simon I. Hay, and Peter W. Gething (2014) 'Air temperature suitability for Plasmodium falciparum malaria transmission in Africa 2000-2012: a high-resolution spatiotemporal prediction.' *Malaria Journal* 13(1), 171