# Disease and Diversity in Africa's Long-term Economic Development

Javier A. Birchenall*
*University of California at Santa Barbara*

July 3, 2011

**Abstract**

This paper establishes a line of continuity between pre-colonial economic conditions and the modern era. Using the Standard Cross-Cultural Sample, a rich ethnographic account of pre-industrial societies at the time of their first encounter with Europeans, I find that pathogen stress has a negative and significant effect on the sophistication of pre-modern economies. A likely channel is the effect of disease on social segmentation and ethnic diversity. A causal link between disease and diversity holds even in current times, especially in Africa. The results weigh against the "colonial origins" hypothesis that disease influences current economic outcomes only through European settlements.

JEL classification: J11; O11; P51

Communications to: Javier A. Birchenall
Department of Economics, 2127 North Hall
University of California, Santa Barbara CA 93106
Phone/fax: (805) 893-5275
E-mail: jabirche@econ.ucsb.edu

*A previous version of this paper circulated with the title "Africa's past economic development and its determinants." I would like to thank Philip Babcock, Sambit Bhattacharyya, Hoyt Bleakley, Bob Fogel, David Carr, Juan Carlos Cordoba, Dave Donaldson, Stanley Engerman, James Fenske, Nicola Gennaioli, Ana Maria Ibáñez, Hyeok Jeong, Michael Kevane, Gary Libecap, Ted Miguel, Kris Mitchener, Nathan Nunn, Pekka Hämäläinen, Phil Hoffman, Jeff Nugent, Bonnie Queen, James Robinson, Jean-Laurent Rosenthal, Raaj Sah, Romain Wacziarg, Warren Whatley, and seminar participants at multiple locations for very useful comments. I would like to specially acknowledge the many suggestions and criticisms by Peter Lindert, my discussant at the All-UC EH meeting.

# 1 Introduction

There are large economic differences in the world, but from the colonial era to the present, economic progress in sub-Saharan Africa has been relatively slow. Although no single factor is likely to have been responsible for Africa's poverty, much of the current literature focuses on modern influences such as the legacies of European colonialism and the slave trade (see, e.g., Acemoglu et al. [2]; Nunn [36]). This emphasis, however, has left out many potential lines of continuity between pre-colonial conditions and the modern era.

This paper presents a comparative study of Africa's long-term development and proposes a line of continuity between pre-colonial and current conditions. Pre-colonial aspects are of considerable interest partly because Africa remains a negative case even after the modern influences emphasized in current discussions are taken into account and partly because understanding fundamental influences that pre-date the European arrival is useful in discriminating between competing explanations of Africa's weak performance.

The major difficulty in undertaking an analysis of pre-colonial conditions is the absence of well-established data. In this paper, I rely on Murdock and White's [35] Standard Cross Cultural Sample (SCCS). The SCCS consists of disaggregate data from 186 ethnographically well-described societies with different subsistence strategies. The SCCS was designed to be representative of all the pre-industrial societies in the world and it was constructed to maximize independence in terms of cultural and historical origin. Societies are thoroughly described from historic and ethnographic literature at the time coinciding with or just after contact with Western cultures when some reliable observer (i.e., traveler, missionary, trader, colonial agent, or anthropologist) first visited the society and wrote about it in sufficient detail (see, e.g., Pryor [38]). The SCCS represents the earliest systematic data one can obtain about tropical Africa and the New World.

The SCCS measures various economic aspects of pre-modern societies such as the degree of complexity in the buildings of a given society. For example, the SCCS reports evidence associated with "large or impressive structures." Physical evidence of these structures is easy to spot by outside observers and hence it is an aspect of pre-modern societies that can be measured with relatively high precision. Historically, impressive structures are a manifestation of the state capacity, or the capacity of urban elites, to extract resources from peasants. This suggests that large structures are a reasonable *proxy* for the presence of large tributary empires and the state's capacity to generate, extract, and mobilize resources in a society.

I find a negative and statistically significant difference in large or impressive structures between African societies and the rest of the world, and even differences with the societies in tropical America, which, in many respects, are quite comparable to African societies. To better understand pre-colonial conditions in Africa, I also follow Acemoglu et al. [2] and study the number of medium- and large-sized cities from archeological data. Chandler's [12] inventory of cities also suggests that sub-Saharan Africa had fewer and less densely populated cities than comparable regions. For example, while there were no large cities south of the Sahara at the time of the European expansion, Teothihuacán (currently Mexico city) was among the ten largest cities of the world in 400 A.D., Chandler ([12], 464). Thus, Westerners' accounts of large or impressive structures and archeological inventories of cities both suggest weak state capacities in pre-colonial Africa. This finding conforms well with many existing discussions, see, e.g., Connah [14]; Curtin et al. [15]; Englebert [19]; Fortes and Evans-Pritchard [21]; Herbst [23]; Hopkins [24].

The previous findings raise a question that I address in the paper, what factors could account for the absence of large tributary empires in pre-colonial Africa? Evidence drawn from the SCCS suggests that pathogen stress, a measure of the prevalence and severity of serious pathogenic diseases, is quantitatively important in the comparative development of pre-industrial societies. I find a robust negative relationship between large or impressive structures and pathogen stress. This negative relationship is driven chiefly (but not exclusively) by African societies which, for a multiplicity of environmental reasons, have faced more adverse disease environments than societies in tropical areas elsewhere.

Several reasons suggest that the negative association between pathogen stress and premodern sophistication is *causal*. As I discuss later on, reverse causality is probably not an important concern. Further, these pathogens are typically transmitted through vectors environmentally determined; they are not by-products of changes in food production, and they have a regional distribution concentrated in tropical Africa.

This finding is important, in part, because the literature that examines modern differences in income per capita assigns a special role to disease environments, e.g., Acemoglu et al. [1]; Bhattacharyya [7]; Bloom and Sachs [10]; Kamarck [25]. One very popular theory, the "colonial origins" view proposed by Acemoglu et al. [1], argues that disease influenced current economic conditions through colonial institutions. A key identifying assumption in their analysis is that diseases influence current outcomes exclusively through their effect on European settlements and institutions. The importance of pathogen stress in the pre-industrial development of African societies, documented here, contradicts their exclusion restriction as it suggests that pathogen stress may have also influenced pre-colonial conditions. These pre-colonial conditions left a legacy which still has a persistent effect on current outcomes. In fact, *settler mortality* as a determinant of current institutions becomes insignificant once pathogen stress is added to their regressions. This result suggests that disease had a strong influence in both pre- and post-colonial institutions, and hence it cannot provide an exclusion restriction to study colonial influences.

The relationship between disease and pre-modern sophistication, however, does not provide much information about the potential mechanisms by which diseases and economic development have interacted in the past. In this paper, I explore a potential link between disease and ethnic diversity: I show that pathogen stress has a strong influence in pre-modern ethnic diversity, and that ethnic diversity has a strong influence on past economic outcomes. Many previous studies have documented the importance of pathogen stress for the reproductive, social, and cultural isolation of human populations, see, e.g., Cashdan [11]; Fincher and Thornhill [20]. I present a more detailed discussion of these points in a later section. At this stage, I simply want to list a couple of notable examples: the tendency for trade to take place at the borders of territories rather than at centralized markets in pre-colonial Africa, Azevado [6], and the emergence of a rigid caste system in India, which, according to McNeill [31], represented an attempt at *social quarantine.* I find strong support for a causal link between disease and ethnic diversity: Africa's high levels of ethnic diversity were chiefly shaped by environmental factors.

In a final step, the importance of disease and diversity for pre-industrial outcomes raises a separate question: can part of Africa's current performance be explained by its pre-colonial history? I find that pathogen stress and pre-industrial ethnic diversity have a dramatic impact on current income per capita, even once multiple controls for colonial influences are taken into account. I also show that pathogen stress has a strong effect on current measures of ethnolinguistic fractionalization. Thus, the analyses presented here tentatively suggest that key social, political, and economic institutions that originated during the pre-colonial era, and once limited economic conditions and state capacity in Africa, withstood European influences with remarkable persistence and still stand in the way of modern economic growth.

The results presented here resonate with many previous studies. The notion that diseases influenced economic and political outcomes in agricultural societies long before the European expansion has been well demonstrated by historical analysis. According to McNeill ([31], 43), "the high exposure to disease, more than anything else, is why Africa remained backward in the development of civilization when compared to temperate lands (or tropical zones like those in the Americas)." The paper also complements a growing body of research that highlights the pre-colonial origins of current weak or strong state capacities, see, e.g., Acemoglu et al. [3]; Bockstette et al. [9]; Englebert [19]; Gennaioli and Rainer [22]; Herbst [23]. A notable feature of my analysis is that I rely on first-hand evidence about pre-industrial societies. To the best of my knowledge, the SCCS has not been used to examine pre-industrial economic disparities. Finally, I provide new insights into the underpinnings of the high levels of ethnic diversity in Africa, see, e.g., Easterly and Levine [18]. I show that high ethnic diversity existed in Africa even during pre-colonial times and that it is not exclusively associated with colonial influences. I also propose a simple and parsimonious link that connects Africa's distinctive disease environment, its diversity, and its long-term economic performance.

The remainder of the paper is as follows. Section 2 discusses the SCCS data. Section 3 presents the main empirical results. Section 4 connects past influences with current economic conditions. Section 5 concludes this paper.

## 2 Data and measurement

**Societal data.** This paper uses the Standard Cross-Cultural Sample (SCCS) developed by Murdock and White [35] during the 1960s. The SCCS includes 186 pre-industrial societies with various subsistence strategies, including hunter-gatherers, fishers, pastoralists, horticulturalists, and agriculturalists. The SCCS provides extensive coded data constructed from historical records and published field research by ethnographers.

The analysis of past conditions in the SCCS has some limitations. Information about these societies is obtained through narratives and descriptions of Westerners in their first encounters with these societies. This implies that there is no *pristine* society in the SCCS. In fact, many of the native societies in Africa were not reached by Europeans until late in the nineteenth century or early in the twentieth century, see Table 1 below. As pointed out by Pryor ([38], 24-25), the key is to "ask how much these preindustrial societies have changed over the millennia until the pinpointed date." I assume that contamination in the SCCS is not a serious problem and that potential biases are corrected by controls for the date of pinpointing of the society. Further, the SCCS is not a random sample of all pre-industrial societies and, by construction, the SCCS is prone to measurement errors.[1] The SCCS, however, has wide geographic coverage and it is the earliest and most complete systematic assessment of pre-industrial societies in the world.

The geographic composition of the 186 societies is as follows: 32 societies are in sub-Saharan Africa, 35 in South and Central America, 24 in West Eurasia, 34 in East Eurasia, 31 in the insular Pacific and 30 in North America.[2] In the sample, there are 131 agricultural societies. In these societies, the contribution of agriculture to the local food supply is at least 10 percent according to the coded variable §v3 in the SCCS. These

[1] The societies in the SCCS are a culturally independent sub-sample of the *Ethnographic Atlas*, also consolidated by George Murdock. The Atlas contains many more societies than the SCCS but there is less information for each society, i.e., there is no measure of physical structures in the Atlas, which is the central outcome of interest in this paper. Later sections examine current economic outcomes.

[2] The geographical distribution of societies is slightly different from the one in the SCCS (§v200) because I treat South and Central America as a single unit. These regions and the West Indies form a single Neotropical region. (North America is part of the Nearctic zone.) Communication between South and Central America was far more common than between Central and North America, see Diamond [17]. I also treat African societies in the Sahel as part of sub-Saharan Africa whereas, in the SCCS, some are seen as part of Eurasia.

societies are also distributed relatively equally among the major regions of the world. The remaining 55 societies are hunter-gatherers. I consider all societies but add a control for hunter-gathering because the economic and political organization of hunter-gatherers is very different from the organization of agricultural societies. (In a working paper version, I simply excluded hunter-gatherers from the sample. The results are virtually identical.)

Table 1. Descriptive statistics for agricultural societies in the SCCS.

| | | | Region | | |
|---|---|---|---|---|---|
| | Mean | Std. Dev. | Africa | America | Eurasia |
| | A. Large or impressive structures (§v66) | | | | |
| Large or impressive structures | 2.40 | 1.38 | 1.50 | 2.03 | 2.82 |
| | B. Demography | | | | |
| Population density | 4.53 | 1.72 | 4.80 | 3.24 | 4.85 |
| Community size | 3.98 | 1.41 | 3.92 | 3.50 | 4.16 |
| | C. Geography and technology | | | | |
| Distance from equator | 18.12 | 13.29 | 9.15 | 13.30 | 22.65 |
| Log-altitude | 4.40 | 2.62 | 5.74 | 4.65 | 3.90 |
| Agricultural potential | 17.66 | 2.55 | 18.42 | 18.30 | 17.20 |
| Food surplus and storage | 1.90 | 0.71 | 1.69 | 1.76 | 2.01 |
| Technological sophistication | 22.36 | 6.39 | 21.88 | 18.34 | 23.84 |
| | D. Disease environments and ethnic diversity | | | | |
| Pathogen stress | 13.59 | 3.31 | 17.46 | 13.73 | 12.27 |
| Ethnic groups within 250 miles | 7.53 | 9.47 | 18.73 | 3.76 | 5.00 |
| Ethnic groups within 500 miles | 23.86 | 26.33 | 59.19 | 11.34 | 16.06 |
| | E. Date of pinpointing of the society (§v838) | | | | |
| Pinpointing date | 1873 | 213.02 | 1915 | 1868 | 1873 |

Note.– The sample is based on 131 agricultural societies in the SCCS. Africa, America, and Eurasia are the means for societies in sub-Saharan Africa, South and Central America, and Eurasia. Many of these variables are ordinal in scale.

In this section I focus on the degree of complexity in the buildings and structures in a given society. I use the coded variable for "large or impressive structures," which is §v66 in the SCCS. I present complementary evidence of alternative *proxies* of pre-modern sophistication in the Appendix, and in a subsequent section I will examine current income per capita. I focus on the presence of large structures because they are one type of *state capacity* historically associated with tributary empires, urban life, and surplus extraction.

Further, large or impressive structures can be measured retrospectively with a high degree of accuracy and subjective assessments are much less important than in other measures in the SCCS.

In the SCCS, large or impressive structures is organized in an *ordinal* scale with higher values representing higher complexity. For agricultural societies, this variable is coded as follows, §v66: 1 = None (51 cases), 2 = Residences of influential individuals (19), 3 = Secular or public buildings (29), 4 = Religious or ceremonial buildings (25), 5 = Military structures (3), 6 = Economic or industrial buildings (4 cases). This variable is coded as 1 in all hunter-gatherers societies. The ordering in the SCCS might not represent higher sophistication so I will also examine a dichotomous variable for the presence of religious, military, or industrial structures only. That is, for cases with §v66>3.

Table 1 reports summary statistics for the agrarian societies. The table shows that large or impressive structures were less prevalent in the Americas compared with Eurasia, and less prevalent in Africa compared with the Americas. The difference recorded by the SCCS is quite likely accurate to the extent that one cannot nowadays point to a large sub-Saharan city in pre-modern times or to evidence of large public works typical of tributary empires. It is clear that African empires existed before Mediterranean and European contact, see, e.g., Davidson [16]; Connah [14]; Hull [26]. Thus, the notion that state consolidation was not possible seems unlikely. Large-scale urbanization in Africa, however, seems less common than in any other agricultural society, especially given the technological sophistication of societies in Africa.

For demographic controls, I will use population density (§v64, defined as number of people per square mile) and community size (§v63). Both variables are ordinal scales from 1 to 7 but they essentially represent the logarithm of density and size respectively. Controls for geography include distance from the Equator and log-altitude (§v183), and a measure of the agricultural potential in the region where the societies are located. Potential for agriculture is an index based on land slope, soil quality, and climate (§v921). I also use an explicit measure of food surplus in the society (§v21). Technological sophistication is a measure associated with the existence of writing and records (§v149), the fixity of residence (§v150), technological specialization (i.e., presence of pottery, metal work, and loom weaving, §v153), land transport (§v154), monetary exchange (§v155), and social stratification (§v158). I add these measures to construct an overall index. This practice is standard in the SCCS.[3]

Later on I will examine ethnic diversity. Measures of ethnic diversity were compiled by Cashdan [11]. These measures are calculated using the number of ethnic groups present within a given radius (100-500 miles in 50-mile increments) of each SCCS society out of a "universe" of ethnographic societies in the world. This measurement procedure differs slightly from indices of ethnolinguistic fractionalization, which measure the probability that two randomly drawn people within a country will be from different ethnic groups. Since societies are the political unit of the SCCS, it is not possible to construct identical measures as those used for countries. (A control for the size of the society, though, serves to align both measures.) However, the SCCS measures and the measures of ethnolinguistic fractionalization from Easterly and Levine [18] are strongly positively related. Further, as discussed in Cashdan [11], SCCS measures of ethnic diversity are highly correlated with alternative estimates of linguistic diversity. My baseline analysis considers measures based on a 500 mile radius (§v1872) although, I also present results with a 250 miles radius (§v1867). As Table 1 shows, ethnic diversity is larger in Africa compared to agricultural societies in any other part of the world.

**Some reduced form results for pathogen stress.** Because I am interested in examining the effects of disease in pre-industrial times, I will rely on a general measure of pathogen stress (§v1260), whose values were obtained from medical and public health sources on the latitude and longitude of the sample societies, using data as close as possible to the defined dates for the sample societies' SCCS data, see Low [27].[4] A total of seven pathogens (leishmanias, trypanosomes, malaria, schistosomes, filariae, spirochetes, and

[3]Africa seems to have experienced an independent origin of iron work often cited as being part of the advancements spread with the Bantu expansions. However, "iron apparently made no dramatic impact upon early African agriculture," Austen ([5], 14). Cattle domestication also seems to have had an independent origin, see Austen ([5], chapter 1). Important independent achievements in writing, mathematics, and science also took place in the New World, see Mann ([29], 16-20 and 63-65).

[4]Low [27] studied disease and marriage systems rather than economic outcomes. Thus, there is no *a priori* sense of bias in these measures, which is an advantage of the indicators of disease in the SCCS.

leprosy) are recorded in §v1253-1259 and are rated on a 3-point scale for frequency and severity. The individual scores are added to yield a total pathogen stress score. A high score represents many types of pathogens and severe exposure. I will later on consider just the presence (regardless of the severity) of the pathogen as a robustness check. In principle, an analysis of individual pathogens is possible, but total pathogen stress is the central variable of interest as there is no reason to predict special associations with particular pathogens. Further, pathogens and their severity tend to be spatially concentrated, and an analysis based on a single disease would fail to recognize these complementarities.

This measure of pathogen stress is plausibly *exogenous*. The pathogens reported by the SCCS are largely unrelated to economic conditions (i.e., sanitation, childhood immunization, etc.) and this implies that reverse causality is not an important concern. Further, these pathogens are typically transmitted through vectors environmentally determined and not through an oral-fecal route which is strongly affected by urbanization and public health interventions. Finally, these diseases are not by-products of changes in food production, as many of the crowd epidemic diseases in temperate areas, and these pathogens have a regional and not a worldwide distribution. For a variety of well-known reasons, tropical diseases have been heavily concentrated in sub-Saharan Africa.[5]

Given that pathogen stress is environmentally determined, the basic empirical strategy is straightforward. I first estimate regression equations of the following form:

$$Y_i = \delta \cdot \text{Pathogen stress}_i + X_i \cdot \beta + \varepsilon_i, \tag{1}$$

where $Y_i$ measures the complexity of large or impressive structures in society $i$, Pathogen stress$_i$ is the measure of pathogen stress in society $i$, and $X_i$ are a series of controls that are meant to capture aspects that are potentially important for the presence and scope of large structures. The coefficient of interest is $\delta$ which measures the reduced form effect

[5] Far more tropical diseases originated and are still found in Africa because humans and Old World monkeys and apes are genetically closer and this implies higher disease transfers, see Wolfe et al. [43]. Further, "the array of parasites that infest wild primate populations is known to be formidable," see McNeill ([31], 15). Monkeys and apes in the Old World also serve as *reservoirs* for many of the human diseases in Africa. This is not the case for New World monkeys. Incidentally, the absence of large mammals in the New World meant reduced opportunities for disease transfers, see Wolfe et al. [43].

of pathogen stress conditional on the controls $X_i$.

Table 2. Reduced form estimates for the impact of disease in pre-industrial societies.

| | Dependent variable: Large or impressive structures | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Pathogen stress | −0.089*** | −0.087*** | −0.079** | −0.102*** | −0.101*** | | −0.076** |
| | (0.023) | (0.023) | (0.030) | (0.029) | (0.029) | | (0.031) |
| Sub-Saharan Africa | | | | | | −0.977*** | −0.503** |
| | | | | | | (0.20) | (0.25) |
| Community size | | 0.21*** | 0.20*** | 0.09 | 0.07 | | 0.08 |
| | | (0.05) | (0.05) | (0.06) | (0.06) | | (0.06) |
| Distance from equator | | | −0.00 | −0.01 | −0.01 | | −0.01 |
| | | | (0.01) | (0.01) | (0.01) | | (0.01) |
| Log-altitude | | | −0.05 | −0.04 | −0.04 | | −0.02 |
| | | | (0.04) | (0.03) | (0.03) | | (0.03) |
| Agricultural potential | | | −0.01 | −0.01 | −0.01 | | −0.01 |
| | | | (0.03) | (0.02) | (0.02) | | (0.02) |
| Food surplus | | | | 0.30** | 0.31** | | 0.30** |
| | | | | (0.13) | (0.13) | | (0.13) |
| Technological sophistication | | | | 0.06*** | 0.06*** | | 0.05*** |
| | | | | (0.02) | (0.01) | | (0.01) |
| Pinpointing date | | | | | −0.00 | | −0.00 |
| | | | | | (0.001) | | (0.001) |
| $R^2$ | 0.18 | 0.24 | 0.24 | 0.32 | 0.32 | 0.20 | 0.34 |
| N. Obs. | 186 | 185 | 181 | 181 | 181 | 186 | 181 |

Note.– All specifications include a (non-reported) control for hunter-gatherers. In parentheses are robust standard errors. ***, ** and * denote statistical significance at the 1, 5, and 10 percent levels, respectively.

Table 2 reports the reduced form estimates. Column (1) includes pathogen stress, not including any covariates. The point estimate is −0.089 (standard error 0.023). Thus, column (1) suggests that large or impressive structures were less complex in agricultural societies affected by heavy pathogen stress. Column (2) controls for demographic influences through the size of the society, since only large societies can afford to devote resources to the construction of large or impressive structures. This variable is statistically significant in some specifications but the size and significance of $\delta$ changes little with the addition of this control. Column (3) controls for geographic differences and this lowers marginally the value of $\delta$. However, there is no statistical difference in point estimates in (1) and (3). In column (4) I include controls for the amount of food surplus in the society and its technological sophistication. These variables are important for large buildings and structures, in all specifications, but the point estimate for pathogen stress actually becomes more negative. In (4), the point estimate for pathogen stress is $-0.102$ (standard error 0.029), which is not distinguishable from the point estimate in (1). Finally, in (5), I add the date of pinpointing of the society to control for potential measurement biases, but this variable has no effect in the estimation.

*Discussion.*– The size of the estimated impact of pathogen stress is large. The difference in pathogen stress between tropical Africa and tropical America is $3.73 = (17.46 - 13.73)$, and the difference between tropical Africa and Eurasia is 5.19, Table 1. Further, the difference in large or impressive structures for these same regions is $-0.53$ and $-1.32$, respectively, Table 1. A point estimate of $-0.09$ indicates that about 60 percent of the difference between Africa and tropical America can be accounted for by differences in pathogen stress. Similarly, about 35 percent of the difference between Eurasia and Africa can be accounted for due to differences in pathogen stress.

Some controls in $X_i$ are endogenous. The fact that the estimates for $\delta$ change little in Table 2 makes the assumption of exogeneity in pathogen stress more credible. Moreover, the fact that controls for food surplus and technological sophistication marginally lower the point estimate for $\delta$ deepens the African puzzle. The presence of food surplus and technological sophistication are central for the organization of pre-modern societies. Despite higher levels of sophistication (and contact with Eurasia), large or impressive structures were less prevalent in tropical Africa compared to tropical America.

Columns (6) and (7) have a separate but related purpose. In both, I have included an African indicator for societies in this region. This approach mimics analyses of the *Africa dummy* in modern growth regressions, see, e.g., Collier and Gunning [13]. Column (6) simply shows the mean difference between Africa and Eurasia. This difference is negative and significant. In (7), I have included all controls and the African indicator variable. The point estimate for pathogen stress declines to $-0.076$ but it is still significant at the 95 percent level. Further, as this column shows, the value of the African indicator is reduced by about one half with these controls. This is consistent with the analysis presented in the previous paragraphs and, in fact, suggests that pathogen stress plays a central role in accounting for Africa's pre-industrial performance.

**Robustness.** Table 3 verifies the robustness of the previous estimates in multiple ways. I use the same specifications of Table 2. Thus, specification (7) includes a sub-Saharan Africa control, which is an important control throughout.

Instead of using the order assumed by the SCCS, in panel I.A, I consider a dichotomous variable for the presence of religious, military, or industrial structures, which are considerably more likely to provide evidence of strong pre-colonial state capacity. In panel I.B, I use a *probit* estimate and the previous dichotomous dependent variable and report the point estimates for $\delta$. In both cases, however, there is a negative and robust association between pathogen stress and this alternative measure of state capacity.

In panel II.A, I consider the existence of the pathogens in the local environment of the society regardless of the severity and endemicity of the disease. The presence of the disease may be considered "more exogenous" than a measure of pathogen stress that combines the presence and the severity of the disease. As this specification shows, the negative relationship between large structures and disease becomes stronger. I also consider in panel II.B the same dependent variable as in Table 2 but use an *order probit* regression instead of a linear regression. The results, once again, are virtually unchanged. In panel II.C, I substitute community size for population density as a demographic control. Further, I add a control for ethnic diversity in panel II.D, and find that there is a negative association between pathogen stress and pre-colonial outcomes in this case as well.

Finally, in panel III, I examine alternative sub-samples. I restrict the sample to tropical areas only. I also exclude the Incas and Aztecs from the sample to increase comparability. I also exclude sub-Saharan African societies in the Sahel. In all these cases, the pattern observed in Table 2 remains unaltered with only minimal changes in the point estimates.

Table 3. Robustness checks for reduced form estimates.

| I. Dependent variable: Religious, military, or industrial structures only | | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (7) |
| | A. OLS estimates | | | | | |
| Pathogen stress | $-0.020^{***}$ | $-0.019^{***}$ | $-0.016^{*}$ | $-0.024^{***}$ | $-0.024^{***}$ | $-0.018^{**}$ |
| | (0.006) | (0.006) | (0.009) | (0.009) | (0.009) | (0.009) |
| | B. Probit estimates | | | | | |
| Pathogen stress | $-0.099^{***}$ | $-0.102^{***}$ | $-0.087^{**}$ | $-0.138^{***}$ | $-0.139^{***}$ | $-0.112^{**}$ |
| | (0.03) | (0.03) | (0.04) | (0.04) | (0.04) | (0.05) |
| N. Obs. | 186 | 185 | 181 | 181 | 181 | 181 |
| II. Dependent variable: Large or impressive structures | | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (7) |
| | A. Using the presence of the pathogen only | | | | | |
| Pathogen presence | $-0.150^{***}$ | $-0.153^{***}$ | $-0.136^{**}$ | $-0.169^{***}$ | $-0.166^{***}$ | $-0.124^{**}$ |
| | (0.04) | (0.04) | (0.05) | (0.05) | (0.05) | (0.06) |
| N. Obs. | 181 | 181 | 181 | 181 | 181 | 181 |
| | B. Order probit estimates | | | | | |
| Pathogen stress | $-0.097^{***}$ | $-0.096^{***}$ | $-0.081^{***}$ | $-0.116^{***}$ | $-0.116^{***}$ | $-0.088^{**}$ |
| | (0.02) | (0.02) | (0.03) | (0.03) | (0.03) | (0.03) |
| | C. Controlling for population density rather than community size | | | | | |
| Pathogen stress | $-0.089^{***}$ | $-0.101^{***}$ | $-0.092^{***}$ | $-0.107^{***}$ | $-0.105^{***}$ | $-0.077^{**}$ |
| | (0.02) | (0.02) | (0.03) | (0.03) | (0.03) | (0.03) |
| | D. Adding controls for ethnic diversity | | | | | |
| Pathogen stress | $-0.089^{***}$ | $-0.086^{***}$ | $-0.087^{***}$ | $-0.101^{***}$ | $-0.096^{***}$ | $-0.075^{**}$ |
| | (0.02) | (0.02) | (0.03) | (0.03) | (0.03) | (0.03) |
| N. Obs. | 181 | 181 | 177 | 177 | 177 | 177 |
| III. Dependent variable: Large or impressive structures. Sub-samples. | | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (7) |
| | A. Sub-Saharan Africa and tropical America only | | | | | |
| Pathogen stress | $-0.060^{*}$ | $-0.085^{**}$ | $-0.080^{**}$ | $-0.080^{**}$ | $-0.07^{**}$ | $-0.05$ |
| | (0.03) | (0.03) | (0.04) | (0.04) | (0.04) | (0.04) |
| N. Obs. | 67 | 67 | 65 | 65 | 65 | 65 |
| | B. Excluding Incas and Aztecs | | | | | |
| Pathogen stress | $-0.086^{***}$ | $-0.085^{***}$ | $-0.071^{**}$ | $-0.094^{***}$ | $-0.094^{***}$ | $-0.071^{**}$ |
| | (0.02) | (0.02) | (0.03) | (0.03) | (0.03) | (0.03) |
| N. Obs. | 184 | 183 | 179 | 179 | 179 | 179 |
| | C. Excluding societies in the Sahel | | | | | |
| Pathogen stress | $-0.085^{***}$ | $-0.103^{***}$ | $-0.107^{***}$ | $-0.112^{***}$ | $-0.110^{***}$ | $-0.083^{**}$ |
| | (0.03) | (0.03) | (0.03) | (0.03) | (0.03) | (0.03) |
| N. Obs. | 128 | 127 | 123 | 123 | 123 | 123 |

Note.– In parentheses are robust standard errors. ***, ** and * denote statistical significance at the 1, 5, and 10 percent level. The specifications coincide with those in Table 2.

**Alternative proxies for pre-modern sophistication.** There are only limited *proxies* for economic conditions in pre-modern times.[6] In the Appendix, I follow Acemoglu et al. [2] and discuss evidence for urbanization rates before 1500 from the well-known archeological estimates of Chandler [12]. (In their empirical analysis of pre-modern urbanization, Acemoglu et al. [2] excluded sub-Saharan Africa although they used data from Chandler [12] for other regions.) I find that cities in sub-Saharan Africa had more recent and external origins, and that they were scarcer and less densely populated than those in tropical America or in North Africa. Urbanization rates were also lower in Africa in pre-colonial times. Thus, if one uses urbanization rates to *proxy* for income per capita, in Acemoglu et al. [2]'s felicitous language, Africa would have experienced a *persistence of misfortunes*. That is, compared to South and Central America, sub-Saharan Africa has been relatively poor even before the European expansion. This finding contrasts with, but does not contradict, the *reversal of fortune* described in Acemoglu et al. [2] among all of the European colonies, e.g., in 1500, North America and Australia were the poorest regions of the world whereas today, both are among the richest nations.[7]

Population size and density are alternative *proxies* commonly used to study pre-modern economies, see Acemoglu et al. [2]. The Appendix reports data from Biraben [8] and McEvedy and Jones [30]. These estimates differ on some dimensions but in both, sub-Saharan Africa has larger populations in 1500 than the Americas, and Africa has the fastest population growth in the world in the years between 400 B.C. (or A.D.) and 1500 (or 1000). Thus, measures of population density give a different view compared to measures based on the existence of large and impressive structures, and measures of urbanization. That is, existing evidence suggests that Africa was more densely populated than tropical America and North Africa. This is also true in Table 1.

[6] Another possibility is *biological standards of living* based on anthropometric measures. However, some African tribes (the Tutsi) are exceptionally tall because this is a genetic adaptation that helps diffuse heat. Others (the Pygmies), are exceptionally short because they lack, due to genetic reasons, the growth spur during adolescence. All this implies that cross-sectional inferences based on height would be uninformative.

[7] In terms of income per capita in 1500, existing estimates also suggest lower income per capita in sub-Saharan Africa compared to the rest of the Old World, and even lower income per capita than in tropical America, see, Maddison ([28], Table 4.1).

This conflicting evidence in terms of pre-modern conditions arises because population density was *not* systematically associated with urbanization and state consolidation in Africa. In all other regions of the world, urbanization rates and population densities are always positively associated, but in Africa they are not. Africa's divergent pattern has been previously documented in multiple independent analyses of political scientists, e.g., Fortes and Evans-Pritchard [21]; Vengroff [41].

Table 4. OLS correlation between large or impressive structures and population density.

| | All societies (1) | Only sub-Saharan Africa (2) | Non-African societies (3) | Eurasian societies (4) | Only America (5) | America without Incas and Aztecs (6) |
|---|---|---|---|---|---|---|
| I. Dependent variable: Large or impressive structures | | | | | | |
| Population density | 0.21*** | −0.29 | 0.27*** | 0.21*** | 0.37*** | 0.30** |
| | (0.05) | (0.23) | (0.05) | (0.06) | (0.12) | (0.15) |
| $R^2$ | 0.19 | 0.14 | 0.29 | 0.27 | 0.35 | 0.29 |
| N. Obs. | 184 | 32 | 152 | 118 | 34 | 32 |
| II. Dependent variable: Religious, military, or industrial structures only | | | | | | |
| Population density | 0.06*** | −0.05 | 0.07*** | 0.06*** | 0.13*** | 0.10** |
| | (0.01) | (0.05) | (0.01) | (0.02) | (0.03) | (0.04) |
| $R^2$ | 0.11 | 0.10 | 0.16 | 0.13 | 0.35 | 0.26 |
| N. Obs. | 184 | 32 | 152 | 118 | 34 | 32 |

Note.– All specifications include a (non-reported) control for hunter-gatherers. In parentheses are robust standard errors. ***, ** and * denote statistical significance at the 1, 5, and 10 percent level.

Africa's divergent pattern can be also seen in the SCCS. Table 4, panel I, shows that there is no systematic association between population density and the presence of large buildings and impressive structures in Africa. Panel II shows the same results for the presence of religious, military, or industrial structures only. These results contrast with a positive association in all other agricultural societies in the world. Specifically, column (1) of Table 4 reports the OLS correlation between the presence of large buildings and population density in all societies. Column (2) examines the same correlation within Africa. This correlation is actually negative although not statistically significant. In non-African regions, this correlation is strong and positive, column (3). This is also true for societies in Eurasia, column (4). In South and Central America, the correlation is actually stronger than in non-African societies, column (5), and not driven by the presence of large empires such as the Incas and the Aztecs, column (6). Acemoglu et al. ([2], Table V) provide additional evidence of a positive relationship between urbanization rates and population density *outside* of Africa.

Acemoglu et al. [2] provide additional reservations regarding the use of population density as a *proxy* for pre-modern conditions. The current empirical link between population density and income per capita is weaker than that of urbanization. For instance, there is no cross-sectional association between population density and income per capita today. In contrast, the association between urbanization and income per capita is strong and robust even today, see Acemoglu et al. ([2], Table II). Further, while the Malthusian model suggests that higher densities may reflect higher income per capita, "the main thrust of Malthus' work was how a higher than equilibrium level of population increases death rates and reduces birth rates to correct itself. A high population density could therefore be reflecting an 'excess' of population, causing low income per capita," see Acemoglu et al. ([2], 1243). Given these reservations, I focus on *proxies* for urban life rather than on measures of population size or density.[8]

# 3 Why would disease influence past outcomes?

The previous section shows that pathogen stress is quantitatively important for understanding pre-industrial outcomes associated with the existence of tributary empires. The distinctiveness of African disease environments has been documented by Arab travelers as well as by early European settlers, see, e.g., Azevado [6]; Curtin et al. ([15], 93). The absence of large tributary empires and widespread state consolidation in Africa is also a conclusion consistent with many previous assessments, see, e.g., Chandler [12]; Connah [14]; Curtin et al. [15]; Herbst [23]; Hull [26]; Modelski [34].

The evidence presented so far, however, does not provide specific channels of causation.

[8] Population size, however, is negatively related to pathogen stress in the SCCS. Outside of the tropics, for example, the point estimate is $-0.10$ (s.e. 0.05).

Pathogens do not seem to be an important *direct* influence on pre-modern outcomes since large urbanization and state consolidation took hold in several temperate areas despite the many crowd epidemic diseases consequence of agriculture. This logic suggests that diseases were likely to affect pre-modern outcomes through an *indirect* channel.

A well-known example of an indirect channel is the influence of disease on European settlements and institutions studied by Acemoglu et al. [1]. The SCCS includes indigenous groups at the time of the European arrival so, in principle, this channel has to be ruled out. In fact, Table 2 suggests, contrary to the exclusion restriction in Acemoglu et al. [1], that the observed influence of disease at colonial or even in current times, may actually pre-date the European expansion.

**Pathogens and ethnic diversity.** In this sub-section I argue, based on historical and anthropological analyses, that high pathogen stress encourages isolation among human populations and that this isolation has lead to high levels of ethnic diversity in Africa. The underlying logic of this mechanism is fairly simple: in settled agricultural societies, the best way to avoid disease is to maintain some form of social segregation. Segregation, in turn, increases the costs of state consolidation as it produces a more diverse and heterogenous social base. Ethnic, linguistic, religious, and cultural diversity are known to have adverse effects of economic outcomes and public policies, see, e.g., Alesina et al. [4]; Easterly and Levine [18]; Miguel and Gugerty [33]; Spolaore and Wacziarg [40].

Segregation due to disease is consistent with the findings of Cashdan [11] and Fincher and Thornhill [20]. These papers studied the geographical patterns of ethnic group distributions and language in the world, and concluded that diversity is shaped chiefly by environmental factors such as unpredictable climate, pathogen stress, and habitat diversity. In particular, Cashdan ([11], 975-976) concludes that "one of the strongest environmental predictors of high ethnic diversity is pathogen stress" and that "this pattern of relationships suggests that pathogens may be an important force in limiting the size of chiefdom/states." Similarly, Fincher and Thornhill ([20], 1289) conclude that "the worldwide distribution of indigenous human language diversity is strongly positively related to human parasite diversity."

Segregation is also consistent with the increased benefits of geographic and economic isolation in high-disease areas. For example, two sociocultural adaptations to disease in Africa were the absence of widespread networks of exchange and the tendency for trade in pre-colonial Africa to be conducted at the border of territories, rather than at centralized markets. Perhaps the best indication of the isolating effect of disease in Africa comes from Azevado ([6], 127) who notes that: "Before the European arrival in the interior of central Africa, trade was localized and organized in such a way that one ethnic group transported its goods to the limits of its district, and the next group did the same." This pattern of exchange was also common in the trans-Saharan trade networks as "aliens encountered severe health problems in West Africa, and merchants tended to turn back at the desert's edge," see Curtin et al. ([15], 93-94). Thus, given the high prevalence of disease and the reduced interaction and communication, it is therefore quite likely that disease translated into less homogeneous populations.

Attempts to integrate and consolidate power during European colonization also created outbreaks of indigenous diseases in Africa. As Azevado ([6], 119) notes, epidemic diseases including *sleeping sickness* (trypanosomiasis) and malaria "accompanied the breakdown of barriers between previously isolated peoples." This evidence further suggests that geographic isolation was beneficial or that increasing intercommunication and long-distance trade (i.e., the use of rivers) were particularly costly in Africa due to disease.

In fact, disease, as an isolating force in agrarian societies, is so common that social *stigma* is discussed many times in the Old Testament. The Old Testament also tells the tale of the *Tower of Babel*, in which language diversity resulted in the lack of large-scale coordination and communication. These discussions are simple representations of the general view proposed here.

Finally, the notion that disease leads to segmentation is consistent with particular case studies, such as India, where social segregation took the form of a rigid caste system, according to McNeill [31]. Outside of sub-Saharan Africa, India is the only area with high levels of ethnolingusitic fractionalization, see Easterly and Levine ([18], Table III). Indian civilization also arose in "a climate analogous to that of the African savanna lands," see

McNeill ([31], 84). In India, "small, self-contained communities of forest-dwelling peoples" also experienced a variety of tropical diseases and parasitic infections, see McNeill ([31], 81-84). Moreover, "instead of digesting the various primitive communities that had occupied southern and eastern India in the manner that was normal north of the Himalayas, Indian civilization expanded by incorporating ex-forest folk as castes, fitting them into the Hindu confederation of cultures as semi-autonomous, functioning entities." In fact, McNeill ([31], 66-67) argues that "the taboos on personal contact across caste lines, and the elaborate rules for bodily purification in case of inadvertent infringement of such taboos, suggests the importance fear of disease probably had in defining a safe distance between the various social groups that became the castes of historic Indian society." The end result was that "the homogenizing process fell short of the 'digestive' pattern characteristic of the other Old World civilizations. Consequently, the cultural uniformity and sociological cohesion of the Indian peoples has remained relatively weak in comparison to the more unitary structures characteristic of the northerly civilizations of Eurasia."

**Empirical estimates.** The following sub-section is organized according to the previous discussion. First, I document a strong positive association between pathogen stress and pre-industrial ethnic diversity. Then, I show that there is a strong negative association between ethnic diversity and past economic outcomes, and examine whether variation in pathogen stress could be used as an *instrumental variable* for ethnic diversity.

*The association between disease and ethnic diversity.*– Figure 1 plots the number of ethnic groups within 500 miles against the SCCS measure of pathogen stress for a selected group of societies. (The Tables below provide many complemetary results.) The figure shows a strong positive relationship between disease and ethnic diversity. There are more distinct ethnic groups in more disease-prone areas.

Table 5 presents OLS estimates of equation (1) using as dependent variable the number of ethnic groups within 500 miles. As in Figure 1, the relationship between pathogen stress and ethnic diversity is strongly positive. Column (1) includes no controls and shows that higher pathogen stress leads to higher ethnic diversity. The magnitude and significance of this effect is robust to the addition of demographic controls, column (2), and

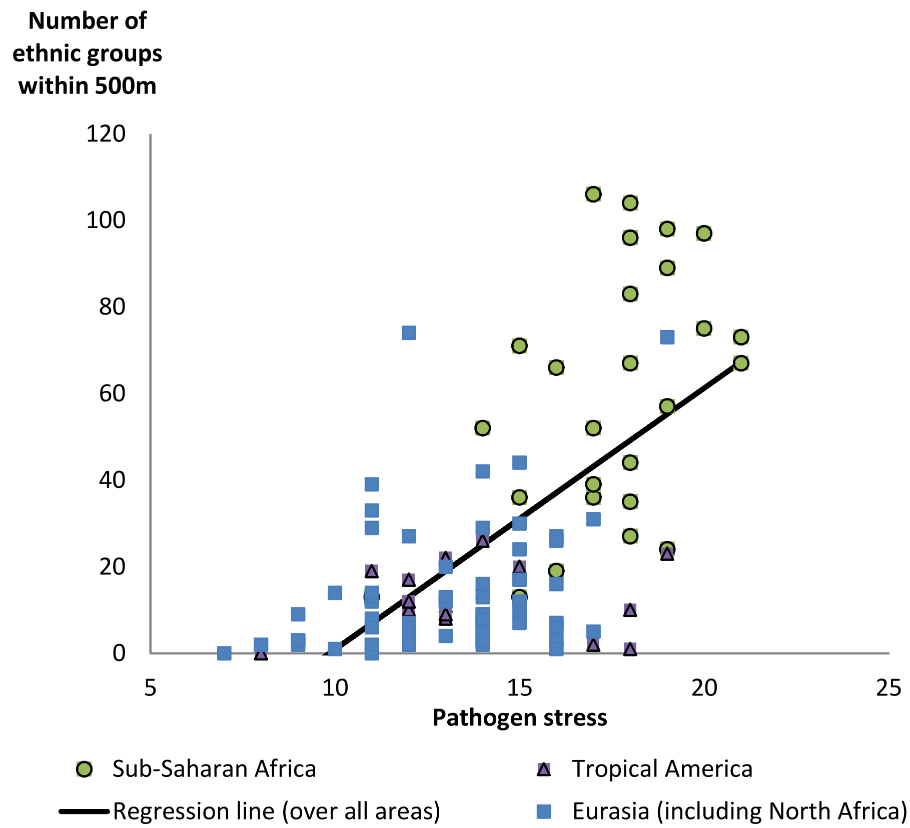


Figure 1: Pathogen stress and ethnic diversity in selected areas. See the text for data details and many robusness checks.

environmental influences such as latitude, altitude, and agricultural potential, column (3). Of these controls, only altitude plays a significant role. A number of studies have suggested that mountains are frequently refuge areas that are difficult to conquer and integrate.[9] The point estimates in (3) survive the inclusion of these controls. Thus, pathogen stress plays a significant role in ethnic diversity beyond the channels of geographic variability studied by the literature. Further, the addition of controls for food surplus and technological sophistication in the society, column (4), leave the previous results virtually unaffected. Higher technological sophistication is negatively associated with ethnic diversity. This aligns well with existing evidence, see Weber [42].

[9] Particular case studies pertaining the Nigerian highlands and the Liangshan mountains between China and Tibet have been referenced by Cashdan ([11], 977). In effect, Michalopolous [32] has proposed a hypothesis to account for modern ethnolinguistic fractionalization emphasizing land quality and elevation.

Table 5. Ethnic diversity and pathogen stress in pre-industrial societies.

| | Dependent variable: Ethnic groups within a 500 miles radius | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Pathogen stress | 3.64*** | 3.68*** | 2.76*** | 3.20*** | 3.17*** | 2.46*** | 0.98 | 1.07 |
| | (0.79) | (0.77) | (0.78) | (0.82) | (0.79) | (0.75) | (0.77) | (0.72) |
| Community size | | 2.10 | 2.52 | 3.62** | 4.24** | 4.87*** | | 3.36** |
| | | (1.38) | (1.50) | (1.64) | (1.72) | (1.69) | | (1.40) |
| Distance from equator | | | −0.32** | −0.17 | −0.14 | −0.17 | | −0.12 |
| | | | (0.15) | (0.18) | (0.17) | (0.17) | | (0.14) |
| Log-altitude | | | 2.59*** | 2.55*** | 2.50*** | 2.09*** | | 1.55** |
| | | | (0.74) | (0.75) | (0.75) | (0.75) | | (0.65) |
| Agricultural potential | | | −0.53 | −0.50 | −0.41 | −0.49 | | −0.94 |
| | | | (0.74) | (0.72) | (0.71) | (0.71) | | (0.61) |
| Food surplus | | | | 1.42 | 1.59 | 2.55 | | 3.21 |
| | | | | (2.89) | (2.83) | (2.81) | | (2.55) |
| Technological sophistication | | | | −0.76** | −0.85** | −0.89** | | −0.88** |
| | | | | (0.36) | (0.36) | (0.36) | | (0.33) |
| Pinpoint date of society | | | | | 0.033* | 0.052*** | | 0.02 |
| | | | | | (0.018) | (0.01) | | (0.01) |
| Slavery | | | | | | 12.91*** | | 0.33 |
| | | | | | | (4.48) | | (3.84) |
| Sub-Saharan Africa | | | | | | | 38.01*** | 34.10*** |
| | | | | | | | (7.17) | (7.46) |
| Tropical America | | | | | | | −6.15** | −9.44** |
| | | | | | | | (2.60) | (3.90) |
| F-statistic | 21.21 | 11.60 | 8.07 | 6.16 | 6.04 | 6.14 | 25.36 | 10.12 |
| $R^2$ | 0.21 | 0.22 | 0.29 | 0.31 | 0.33 | 0.37 | 0.47 | 0.56 |
| N. Obs. | 128 | 127 | 123 | 123 | 123 | 123 | 128 | 123 |

Note.– In parentheses are robust standard errors. ***, ** and * denote statistical significance at the 1, 5, and 10 percent level.

Finally, column (5) adds the date of the pinpointing of the society. This variable is significant at 90 percent confidence but even in this specification, the coefficient on pathogen stress is significant at 99 percent confidence. In fact, the size of the coefficient is about the same as in column (1) with no controls. Column (6) controls for the presence of slavery in the society, which might be an important determinant of ethnic diversity.[10]

[10] A detailed analysis of slavery is beyond the scope of this paper but it has been carried out by Patterson [37] for some societies in the SCCS. Patterson [37] coded the presence and approximate origin of slave populations in the world. I use §v917 in the SCCS, as a control. Separately, Bhattacharyya [7] has shown

For example, focusing on the modern slave trades, Nunn ([36], 164) noted that "slave trades tended to weaken the links between villages, thus discouraging the formation of larger communities and broader ethnic identities." I use the presence of slavery, as coded by Patterson [37]. This specification yields the lowest value of pathogen stress in Table 3, 2.46. This value, however, is still significant at 99 percent confidence levels.

Specifications (7) and (8) have a separate purpose. They add regional controls for sub-Saharan Africa and tropical America to specifications (1) and (6) respectively. In (7) and (8), pathogen stress is insignificant but there is a clear reason for this. Since high ethnic diversity is highly concentrated in sub-Saharan Africa, controlling for this region captures most of the variation in this variable. Notice also that the $R^2$ in (7) is very high. Almost 50 percent of the variation in ethnic diversity is accounted for by these controls. As (8) shows, additional controls yield minor gains in this measure. The regional control for Africa only includes societies south of the Sahara. As the robustness checks show, including all African societies sometimes yields statistically significant results.

*The association between ethnic diversity and large structures.–* Table 6, panel A, presents OLS correlations between large or impressive structures and ethnic diversity, using as controls the variables specified in Table 5. Table 6 shows that there is a strong negative association between ethnic diversity and the presence of large and impressive structures. If causal, this evidence would suggest that the high levels of ethnic diversity in Africa created an important barrier for state consolidation (as *proxied* by the presence of large structures) in pre-colonial times. However, ethnic diversity is an endogenous variable. As stressed by many modern growth analyses, linguistic, cultural, and ethnic fragmentation tends to disappear after the consolidation of states, see, e.g., Weber [42]. Thus, the OLS correlations in Table 6 may simply reflect *reverse causality.* For example, Africa's high ethnic diversity may be a consequence, rather than a cause, of weak state capacity in pre-colonial times.

*Can disease be used as an instrument for ethnic diversity?–* A significant component of ethnic diversity is environmentally determined, as shown in Table 5. Thus, pathogen

that some of the inferences in Nunn [36] do not survive the inclusion of controls for disease. This finding strengthens the arguments discussed here.

stress may be used as an *instrumental variable* for ethnic diversity in order to estimate the effect of ethnic diversity on the complexity of large or impressive structures. The validity of this strategy rests on two assumptions. First, pathogen stress needs to be a source of exogenous variation in ethnic diversity, which is a point documented in Table 5 and the previous sub-section. As in the empirical approach pioneered by Acemoglu et al. [1], it is not necessary for pathogen stress to be the only or the main cause of variation in ethnic diversity.

Table 6. Large or impressive structures and ethnic diversity in pre-industrial societies.

| | Dependent variable: Large or impressive structures | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | A. OLS correlations | | | | | |
| Ethnic groups | $-0.015^{***}$ | $-0.016^{***}$ | $-0.013^{***}$ | $-0.013^{***}$ | $-0.012^{***}$ | $-0.010^{***}$ |
| within 500 m. | (0.003) | (0.003) | (0.003) | (0.003) | (0.003) | (0.004) |
| $R^2$ | 0.08 | 0.18 | 0.16 | 0.24 | 0.25 | 0.26 |
| | B. IV, second-stage, using pathogen stress as instrument | | | | | |
| Ethnic groups | $-0.028^{***}$ | $-0.026^{***}$ | $-0.031^{**}$ | $-0.037^{***}$ | $-0.037^{***}$ | $-0.043^{**}$ |
| within 500 m. | (0.008) | (0.008) | (0.014) | (0.012) | (0.012) | (0.019) |
| N. Obs. | 128 | 127 | 123 | 123 | 123 | 123 |

Note.– In parentheses are robust standard errors. ***, ** and * denote statistical significance at the 1, 5, and 10 percent level. The specifications follow those of Table 5.

Second, conditional on the controls included in the regression, pathogen stress should influence large or impressive structures only through its effects on ethnic diversity. This *exclusion restriction* is strong and unlikely to hold in the data because there are multiple unobserved influences and omitted variables. It is impossible to test for exogeneity but in order to minimize concerns with the failure of exogeneity I will add a series of controls and later on I will examine separate sub-samples.[11] However, the causal estimates of ethnic

[11] Table 6 controls for population size, agricultural surpluses, and technological sophistication, which are variables pathogens might influence directly, see, e.g., Kamarck [25]. Further, influences unaccounted for may invalidate the IV strategy. A concern of this sort is military control and direct costs of extending the authority of any polity over space. Military campaigns are difficult in high disease environments because large armies are susceptible to outbreaks of infections disease, see Curtin et al. [15] and Acemoglu et al. [1]. The influence of disease on military control does not seem important. McNeill ([31], 72) argues that "the ravages of these [epidemic] diseases clearly were not enough to disrupt armies regularly not

diversity are likely to be biased. (A weak instrument is unlikely as the F-statistics in Table 5, panel A, are large.) Thus, I will rely on IV as a complement rather than as an dominant alternative over the OLS correlations.

Table 6, panel B, presents the second-stage results. (The first-stage results are in Table 5.) Both IV and OLS point estimates are negative, significant, and predict a large influence of ethnic diversity on past economic outcomes. The IV estimates, however, are larger in absolute value than the OLS correlations, and they become more negative as controls are added. Consider the mean number of ethnic groups in sub-Saharan Africa and Eurasia. These values are 59 and 16 respectively, see Table 1. Using the IV estimates of columns (1) or (2) in Table 6 yields a difference in the large or impressive structures between Africa and Eurasia of $-0.028 \times (59 - 16) = -1.20$. This value is quite close to the difference in this outcome variable, $-1.32 = (1.50 - 2.82)$, see Table 1. Thus, the high levels of ethnic diversity in Africa could account for 90 percent of Africa's pre-industrial differential with Eurasia. The OLS point estimates in (1)-(6) suggest that ethnic diversity would account for about half of the difference between these regions. The point estimates for IV in (3)-(6) yield a larger prediction, but this is likely to be biased.

**Robustness.** I verify the robustness of previous estimates in several ways. My main focus is on the relationship between pathogen stress and ethnic diversity. Thus, I report the IV estimates in the Appendix. As I acknowledged, IV estimates are likely to be biased due to the possible violation of exogeneity.

Table 7 shows that the relationship between pathogen stress and ethnic diversity is robust to the addition of a number of controls and alternative specifications. In panel A, I use the number of ethnic groups within a 250 miles radius. The results are similarly positive and significant. In panel B, I use the presence of the pathogens regardless of the severity, and in panel C, I use population density instead of size as a demographic control. The results in B and C are actually stronger than in Table 5.

to keep population below levels necessary for empire-building." Physiological adaptations, including the sickle-cell trait which enhances resistance to malaria, have been used by previous studies to argue against direct effects of pathogen stress on productive capacity. Particularly, Acemoglu et al. ([1], 1371) argued that "[diseases] had limited effect on indigenous adults who had developed various types of immunities."

Table 7. Ethnic diversity and pathogen stress in pre-industrial societies.

| | Dependent variable: Ethnic groups within a given area | | | | |
|---|---|---|---|---|---|
| | (2) | (3) | (4) | (5) | (6) |
| | A. Using ethnic groups within 250 m. | | | | |
| Pathogen stress | 1.226*** | 0.880*** | 1.057*** | 1.049*** | 0.805*** |
| | (0.22) | (0.27) | (0.28) | (0.28) | (0.29) |
| N. Obs. | 127 | 123 | 123 | 123 | 123 |
| | B. Using the presence of the pathogen only | | | | |
| Pathogen presence | 5.826*** | 3.660*** | 4.371*** | 4.264*** | 2.993** |
| | (1.40) | (1.37) | (1.44) | (1.43) | (1.44) |
| N. Obs. | 127 | 123 | 123 | 123 | 123 |
| | C. Controlling for density instead of population size | | | | |
| Pathogen stress | 3.741*** | 2.881*** | 3.236*** | 3.213*** | 2.579*** |
| | (0.63) | (0.75) | (0.77) | (0.78) | (0.81) |
| N. Obs. | 126 | 122 | 122 | 122 | 122 |
| | D. Societies in sub-Saharan Africa and tropical America | | | | |
| Pathogen stress | 6.332*** | 5.335*** | 4.970*** | 4.980*** | 3.272*** |
| | (1.19) | (1.42) | (1.49) | (1.45) | (1.56) |
| N. Obs. | 52 | 50 | 50 | 50 | 50 |
| | E. Societies in sub-Saharan Africa and Eurasia | | | | |
| Pathogen stress | 4.250*** | 2.822*** | 3.782*** | 3.772*** | 3.321*** |
| | (0.68) | (0.84) | (0.86) | (0.86) | (0.92) |
| N. Obs. | 101 | 97 | 97 | 97 | 97 |
| | F. Excluding societies in the Sahel | | | | |
| Pathogen stress | 4.250*** | 3.619*** | 3.606*** | 3.582*** | 2.842*** |
| | (0.55) | (0.69) | (0.69) | (0.69) | (0.66) |
| N. Obs. | 122 | 118 | 118 | 118 | 118 |
| | G. Societies in Africa (North and sub-Saharan) | | | | |
| Pathogen stress | 5.850*** | 1.944 | 2.273 | 2.308 | 3.127 |
| | (1.69) | (1.76) | (1.82) | (1.87) | (2.08) |
| N. Obs. | 35 | 33 | 33 | 33 | 24 |
| | H. Excluding societies in sub-Saharan Africa | | | | |
| Pathogen stress | −0.052 | −0.276 | 0.069 | 0.106 | 0.121 |
| | (0.57) | (0.62) | (0.65) | (0.65) | (0.65) |
| N. Obs. | 101 | 99 | 99 | 99 | 99 |

Note.– In parentheses are robust standard errors. ***, ** and * denote statistical significance at the 1, 5, and 10 percent level. The specifications follow those of Table 5.

Table 7 also explores different sub-samples. In panel D, I restrict the sample to societies in sub-Saharan Africa and tropical America. This comparison has many advantages since these areas are similar in geographic conditions and resource endowments. Further, the technological specialization in sub-Saharan Africa was not lower (and perhaps even higher) than in tropical America, Table 1. Africa had contact with Eurasia through the Nile River, through the Sahara (by the Arab trade that started during the seventh century A.D.), and through the Indian ocean (by the East African trade in medieval times).[12] The point estimates in this comparison are more positive than for any other sample. This evidence thus highlights the existence of sharp differences in disease environments and ethnic diversity even within fairly similar tropical areas. In panel E, I exclude societies in the New World. In panel F, I exclude societies in the Sahel. The positive association between pathogen stress and ethnic diversity is maintained, and sometimes even increased.

Finally, panels G and H consider societies within and outside of Africa. There is a weak association between pathogen stress and ethnic diversity within African societies in column (2), but the addition of environmental controls render these estimates insignificant. Similarly, panel H shows that in non-African societies there is no association between pathogen stress and ethnic diversity. Thus, while pathogen stress has a negative effect on large or impressive structures even outside of Africa (Tables 2 and 3, column (7)), the potential mechanisms I have studied here, do not generalize to areas with low disease prevalence or to variation within Africa.

# 4 Are pre-industrial influences persistent?

In this section, I present four main results. First, I show that pathogen stress has a negative reduced-form effect on current income per capita. Second, I show that pre-industrial ethnic diversity, as measured in the SCCS, is also negatively associated with current economic conditions. Third, I show that pathogen stress is a strong predictor of current ethnolinguistic fractionalization. Finally, I show that pathogen stress renders

[12] In fact, contact was perhaps enough to expose Africa to epidemiological conditions of Eurasia. It is known that "when African slaves began to come to the new world after 1500, they suffered no spectacular die-off from contact with European diseases, which is sufficient demonstration that in their African habitat some exposure to the standard childhood diseases of civilization must have occurred," see McNeill ([31], 130). Even in the more distant regions, "the Bantu-speaking farmers of southern Africa had some immunities to help protect them from the European diseases that exterminated the indigenous people of Tasmania and decimated those of Patagonia," see Curtin et al. ([15], 242).

settler mortality insignificant as a predictor of current institutional differences in the analysis of Acemoglu et al. [1].

**Pathogen stress and current income per capita.** Because many questions may be raised about the reliability and significance of "large or impressive structures" as a *proxy* for past economic outcomes, I examine current differences in income per capita.

In order to link the SCCS with current income per capita, I match each of the 186 societies to the countries where these societies located. The Appendix lists the countries on which each society resided according to the latitude and longitude provided by the SCCS. (To ensure a correct imputation, I used the Human Relations Area Files, which separately lists the location of different cultures around the world.) The matching of the SCCS to countries is fairly straightforward for agricultural societies because there is limited geographic mobility in these societies. The only minor concern is that, sometimes, many societies located in one country, and sometimes one society located in many countries. I allow for both possibilities and average SCCS variables across countries.

Further, some societies, notably China, have been politically independent and unified for centuries. For these societies, there is only a subtle distinction between pre-industrial societies and modern states. Other societies, notably the empires in the New World, were disintegrated by European colonization. Thus, the use of modern income per capita as a dependent variable requires special care due to colonization.

I control for colonial influences in three main ways. First, I focus on former colonies only. This essentially restricts the sample to countries in Africa and the New World, which has been an important focus here. Second, I control for population densities in 1500 since they were arguably an important determinant of colonial policies. Less densely populated colonies provided fewer incentives for *extractive institutions*, see Acemoglu et al. [2]. Finally, I control for whether the country was a British, French, German, Spanish, Italian, Dutch, or Portuguese colony.

Table 8 presents three sets of results. Panel A examines reduced form estimates between the log of GDP per capita (PPP) in 1995 and pathogen stress. The reduced form estimates are always negative and statistically significant, even once I control for colonial influences, columns (2) to (4). These reduced form estimates show that pathogen stress and current income per capita are negatively associated. This negative point estimate is consistent with the findings for large or impressive structures in Tables 2 and 3, and as I discussed in the previous section, it can be considered as *causal* due to the nature of pathogen stress. Column (5) controls for colonial influences and religion, and column (6) adds controls for natural resources, including whether a country is landlocked or an island. Finally, column (7) adds a variety of environmental and geographic controls. I include the absolute value of latitude of the country, a number of humidity and temperature controls, and controls for soil quality. The number of controls in these specifications is large.[13]

Panel B presents the OLS correlation between ethnic diversity, as measured by the SCCS for pre-industrial societies, and current income per capita. The correlation in specifications (1)-(5) is negative and statistically significant. In columns (6) and (7) the correlation is negative but not statistically different from zero. The additional control variables included in these specifications are associated with differences in endowments and environmental conditions. An important component of ethnic diversity is environmentally determined and (7) has a large number of controls relative to the number of observations. Thus, including temperature, humidity, latitude and similar variables render ethnic diversity insignificant.

Panel C presents the IV estimates under the causal mechanism previously stated. (The first-stage results are in Table 9, below.) As in the case of pre-industrial influences, these point estimates are likely to be biased due to the failure of exogeneity. The IV results are negative, statistically significant, and the point estimates are larger than the OLS correlations in absolute value. Since exogeneity may also fail in these specifications, I do

[13] Unless otherwise noted, I will rely on the variables coded by Acemoglu et al. ([2], Appendix 2) since I want to minimize explicitly any interference introduced by including multiple variables from the SCCS. I obtained all the data from their online Appendix. Religion in (5) includes controls for catholic, protestant, muslim, and other. Mineral reserves in specification (6) include coal production, and measures of natural reserves as percent of the world for gold, iron, silver, zinc, and oil. In (7), I control for average temperature, minimum monthly high, maximum monthly high, minimum monthly low, and maximum monthly low. Humidity variables include morning minimum, morning maximum, afternoon minimum, and afternoon maximum. Soil quality includes steppe (low latitude), steppe (middle latitude), desert (middle latitude), dry steppe wasteland, desert dry winter, and highland. In total, specification (7) has 25 controls including seven colonial identities, population density in 1500, and 17 environmental controls.

not emphasize these differences.

Table 8. Income per capita in former European colonies, disease, and ethnic diversity.

| | Dependent variable: log-GDP per capita | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| | A. OLS correlations | | | | | | |
| Pathogen | $-0.219^{***}$ | $-0.161^{***}$ | $-0.188^{***}$ | $-0.130^{***}$ | $-0.135^{***}$ | $-0.104^{**}$ | $-0.096^{**}$ |
| stress | (0.026) | (0.027) | (0.034) | (0.033) | (0.033) | (0.039) | (0.040) |
| $R^2$ | 0.43 | 0.54 | 0.52 | 0.63 | 0.65 | 0.71 | 0.78 |
| | B. OLS correlations | | | | | | |
| Ethnic groups | $-0.019^{***}$ | $-0.014^{***}$ | $-0.013^{***}$ | $-0.009^{**}$ | $-0.010^{**}$ | $-0.005$ | $-0.003$ |
| within 500 m. | (0.003) | (0.003) | (0.014) | (0.004) | (0.004) | (0.004) | (0.004) |
| $R^2$ | 0.23 | 0.47 | 0.36 | 0.57 | 0.59 | 0.68 | 0.76 |
| | C. IV, second-stage, using pathogen stress as instrument | | | | | | |
| Ethnic groups | $-0.042^{***}$ | $-0.030^{***}$ | $-0.044^{***}$ | $-0.029^{***}$ | $-0.032^{***}$ | $-0.023^{**}$ | $-0.024^{*}$ |
| within 500 m. | (0.010) | (0.007) | (0.014) | (0.010) | (0.011) | (0.011) | (0.013) |
| Controls included | | | | | | | |
| Pop. density 1500 | No | Yes | No | Yes | Yes | Yes | Yes |
| Colonizer id. | No | No | Yes | Yes | Yes | Yes | Yes |
| Religion | No | No | No | No | Yes | No | No |
| Nat. resources | No | No | No | No | No | Yes | No |
| Lat. and climate | No | No | No | No | No | No | Yes |
| N. Obs. | 67 | 66 | 66 | 66 | 65 | 66 | 66 |

Note.– In parentheses are robust standard errors. $^{***}$, $^{**}$ and $^{*}$ denote statistical significance at the 1, 5, and 10 percent level. For variable definitions please see the main text.

*Discussion.*– Table 8 shows that income per capita is negatively associated with pathogen stress. Given the results of previous sections, this negative association is unlikely to be due to modern influences alone.

To assess the quantitative importance of pathogen stress, recall that the difference in pathogen stress in sub-Saharan Africa and tropical America is 3, which changes little once the societies have been matched to actual countries. The log-income gap between both regions in the sample is $-1.05 = \ln(\$4,478) - \ln(\$1,564)$. Thus, using the OLS point estimate of panel A, column (7), gives a predicted difference of $-0.096 \times 3 = -0.28$, which is about 25 percent of the actual difference between both regions. The point estimates in column (4), which control for colonial influences, explain about 40 percent of the actual difference between both regions.

The previous values are large especially since I have already considered some competing explanations. For example, as I have controlled for colonizer identities and population density in 1500, the estimates already take into account potential colonial influences associated with the *reversal of fortune* studied by Acemoglu et al. [2].[14]

**Pathogen stress and ethnolinguistic fractionalization.** I next substitute the SCCS measures of ethnic diversity by the average of five different indices of *ethnolinguistic fractionalization*, ELF, from Easterly and Levine [18], as used in Acemoglu et al. [1]. This allows me to minimize even more the concerns one may have about the SCCS data or its matching to current countries. In addition, since ethnic diversity is not a fixed category over time, using ELF also serves to examine potentially more modern measures of diversity. (The correlation between the SCCS measure of ethnic diversity and ELF, however, is above 60 percent.)

Table 9 presents the results. The table shows three things: first, in panel I.A, the influence of pathogens stress on ELF is positive and statistically significant, with the exception of (7). In all columns but (7) the point estimate is stable. Thus, Table 9 suggests that there are important environmental determinants of ethnic diversity which were not drastically perturbed by colonial influences. This point is important because the literature has focused on colonization as a factor that created more ethically diverse societies in Africa. In specification (7), the OLS correlation between pathogen stress and ELF is not different from zero. As I stressed before, since ethnic diversity seems to be, in a large part, environmentally determined, pathogen stress becomes less relevant as an explanation for ethnic diversity once temperature, humidity, soil quality, and latitude are controlled for.

The second result in Table 9, panel II.A, previously documented by Easterly and Levine [18], is the strongly negative OLS correlation between ELF and income per capita.

[14] I cannot directly control for urbanization which is a more precise measure of pre-modern conditions. Adding large or impressive structures as a control leaves the results intact. The Appendix discusses urbanization and the implications of pre-colonial differences in this variable.

Finally, panel II.B presents IV estimates for ELF. The point estimates under IV are more negative than the OLS estimates. As the IV strategy may yield biased estimates, I do not give special emphasis to these results.

Table 9. Ethnolinguistic fractionalization, pathogen stress, and income per capita.

| I. Dependent variable: Ethnolingusitic fractionalization | | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| | A. OLS correlations | | | | | | |
| Pathogen stress | 0.036*** | 0.037*** | 0.028*** | 0.030*** | 0.032*** | 0.032** | −0.005 |
| | (0.011) | (0.012) | (0.009) | (0.011) | (0.011) | (0.015) | (0.017) |
| $R^2$ | 0.13 | 0.14 | 0.33 | 0.34 | 0.41 | 0.53 | 0.59 |
| II. Dependent variable: log-GDP per capita | | | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| | A. OLS correlations | | | | | | |
| Ethnolingusitic fractionalization | −1.74*** | −1.47*** | −1.47*** | −1.25*** | −1.30*** | −0.89*** | −0.77** |
| | (0.31) | (0.32) | (0.48) | (0.36) | (0.40) | (0.36) | (0.37) |
| $R^2$ | 0.25 | 0.53 | 0.42 | 0.63 | 0.64 | 0.71 | 0.80 |
| | B. IV, second-stage, using pathogen stress as instrument | | | | | | |
| Ethnolingusitic fractionalization | −6.03*** | −4.29*** | −6.30*** | −4.13** | −4.22*** | −3.17* | −29.15 |
| | (1.58) | (1.26) | (2.07) | (1.59) | (1.45) | (1.78) | (162.32) |
| Controls included | | | | | | | |
| Pop. density 1500 | No | Yes | No | Yes | Yes | Yes | Yes |
| Colonizer id. | No | No | Yes | Yes | Yes | Yes | Yes |
| Religion | No | No | No | No | Yes | No | No |
| Nat. resources | No | No | No | No | No | Yes | No |
| Lat. and climate | No | No | No | No | No | No | Yes |
| N. Obs. | 66 | 65 | 66 | 65 | 65 | 65 | 65 |

Note.– In parentheses are robust standard errors. ***, ** and * denote statistical significance at the 1, 5, and 10 percent level. For variable definitions please see the main text.

**Robustness**. For robustness, the Appendix reproduces Tables 8 and 9 for alternative specifications. I use the presence of the previously listed pathogens discounting their severity and endemicity. I also add controls for the food surplus and technological sophistication based on the SCCS data. I use measures of pre-industrial ethnic diversity based on the number of ethnic groups within a 250 miles radius. The point estimates change across specifications in the same way as in the main tables, but all specifications yield negative and significant results. In fact, the presence of pathogens yields stronger results compared to the previous tables.

I also expand the sample to include non-colonies, and find that colonial influences, measured by the colonizer identity and the population density in 1500, play no role in the point estimate for ethnic diversity once non-colonies are included. This suggests that the colonial controls I have used are indeed associated with potential influences due to colonization. I also find that the influence of ethnic diversity at pre-industrial times in non-colonies is more negative than in the case of only colonies. This suggests that past influences, such as a homogenous population during pre-industrial times, are important for current economic outcomes, more so for countries not colonized by Europeans. (See Spolaore and Wacziarg [40] for a related discussion.)

In other checks, I excluded sub-Saharan Africa from the sample. None of the results are significant, which is consistent with the discussion of the previous section. Sub-Saharan Africa is the most important source of variation in ethnic diversity or ethnolinguistic fractionalization in the data. Thus, excluding this region renders all estimates insignificant. However, even a contrast with North Africa yields strong results. Finally, I include countries in tropical America and Africa only. In this case, the point estimates are also negative and significant.

**Overview.** The findings that deserve special mention are the following. First, pathogen stress influenced negatively pre-industrial conditions. This influence, plausibly causal, is robust and it holds even outside of sub-Saharan Africa. Second, a likely mechanism by which disease influenced past conditions is through ethnic diversity, although high ethnic diversity is an African phenomena and its importance does not generalize to all areas of the world. Third, the previous assessments apply as well to the analysis of current income per capita and ethnolinguistic fractionalization. These insights remain valid even once colonial influences, as measured by the colonizer identity and the population density in 1500, are taken into account. In short, these findings suggest a *pre-colonial origin* for current income differences, see, e.g., Acemoglu et al. [3]; Bockstette et al. [9]; Englebert [19]; Gennaioli and Rainer [22]; Herbst [23], for related discussions.

**Pathogen stress and "colonial origins."** In this final sub-section I briefly revisit the "colonial origins" hypothesis proposed by Acemoglu et al. [1], in which the authors consider a separate causal path through which disease environments could cause low income per capita. The key assumption in their analysis is that diseases, as measured by the mortality rates of potential European settlers, influence current outcomes through their effect on European settlements and institutions alone.

Table 10. Protection against expropriation risk, settler mortality, and pathogen stress.

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Dependent variable: Protection against expropriation risk | | | | | |
| Pathogen stress | | | $-0.361^{***}$ | $-0.327^{***}$ | $-0.257^{***}$ | $-0.242^{***}$ |
| | | | (0.046) | (0.059) | (0.068) | (0.069) |
| Settler mortality | $-0.646^{***}$ | $-0.654^{***}$ | | | $-0.238$ | $-0.234$ |
| | (0.125) | (0.152) | | | (0.179) | (0.197) |
| Linked to SCCS | No | Yes | Yes | Yes | Yes | Yes |
| Country in AJR | Yes | Yes | No | Yes | Yes | Yes |
| Only colonies | No | No | No | No | No | Yes |
| $R^2$ | 0.30 | 0.29 | 0.42 | 0.40 | 0.42 | 0.38 |
| N. Obs. | 74 | 58 | 81 | 58 | 58 | 54 |

Note.– In parentheses are robust standard errors. ***, ** and * denote statistical significance at the 1, 5, and 10 percent level. Linked to SCCS means that at least one society in the SCCS resided in the country. Country in AJR means that the country is part of the sample studied by Acemoglu et al. [1]. Only colonies means that the country is a former European colony.

Since pathogen stress influenced pre-industrial societies, their exclusion restriction may not necessarily hold. Furthermore, settler mortality becomes insignificant as a predictor of institutions (e.g., the protection against expropriation risk) once pathogen stress is included as a control. In Table 10, I have estimated the relationship between settler mortality and protection against expropriation risk including pathogen stress. (This relationship corresponds to the first stage in Acemoglu et al. [1].) Column (1) includes the estimate of settler mortality on the baseline sample of Acemoglu et al. [1]. Column (2) presents the same estimate for the 58 countries with data on settler mortality also linked to the SCCS. The point estimate in both columns is virtually identical and consistent with the point estimates in Acemoglu et al. ([1], Tables 3-6). Both values suggest that disease loads have caused less protective institutions. Columns (3) and (4) include pathogen stress on similar sets of countries. As in the case of columns (1) and (2), the estimates in (3) and (4) are remarkably close. This implies that a sample where both pathogen stress and settler mortality are available is representative of the entire set of countries considered here and in Acemoglu et al. [1].

Column (5) includes pathogen stress and settler mortality as predictors of protection against expropriation risk. There, the value of settler mortality becomes insignificant. The point estimate for pathogen stress, on the other hand, is reduced from $-0.36$ to $-0.25$ but it is significant at 99 percent confidence. Column (6) includes both disease measures but restricts the sample to former colonies with very small changes in the point estimates. Both of these results imply that pathogen stress was an important component behind the mortality faced by European settlers. Table 10, however, implies that an IV strategy is unable to identify the importance of causal channels associated with settler mortality. The problem is that disease played a central role in the organization of pre-colonial societies but also in the selective penetration of European settlers. Thus, disease environments are unable to shed light on the causal effect of colonial influences.

# 5 Conclusions

This paper studied the influence of disease on long-term comparative economic development. I used the Standard Cross-Cultural Sample, a rich ethnographic account of pre-industrial societies described at the time of contact with Europeans. The SCCS reports the complexity of large or impressive structures, which I have used to *proxy* for pre-modern sophistication and the state's capacity to generate, extract, and mobilize resources in a society. I found a robust negative relationship between pathogen stress and this outcome measure. This relationship shows that the comparative development of pre-industrial societies depended critically on their disease environments.

The impact of disease on economic conditions prior to the European expansion challenges common causal mechanisms which explain current income differences across countries by the effect of disease environments in European settlements and institutions, e.g., Acemoglu et al. [1]. Instead, this finding suggests a line of continuity from pre-colonial times. I argued, based on historical analysis, that disease was a key shaper of the organization of pre-colonial societies. Thus, one likely consequence of the high disease prevalence in Africa has been its fragmentation and high ethnic diversity, see, e.g., Azevado [6]; Cashdan [11]; Fincher and Thornhill [20]; McNeill [31].

Pathogen stress has a large and significant impact on current income per capita and ethnolinguistic fractionalization. Both of these results are robust to many relevant controls, such as controls for colonial influences, but they tend to be concentrated in African countries (especially the influence of disease on ethnic fragmentation), and they do not generalize to all areas of the world. These findings tentatively suggest that Africa's economic stagnation has deep roots in its pre-colonial past.

# References


[1] Acemoglu, D., Johnson, S., and Robinson, J.A. (2001) "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review*, Vol. 91, 1369-1401.

[2] Acemoglu, D., Johnson, S., and Robinson, J.A. (2002) "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distributions," *Quarterly Journal of Economics*, Vol. 107, 1231-1294.

[3] Acemoglu, D., Johnson, S., and Robinson, J.A. (2003) "An African Success Story: Botswana," in Rodrik D. (Ed.) **In Search of Prosperity: Analytic Narratives on Economic Growth**, Princeton University Press, 80-122.

[4] Alesina, A., Devleeschauwer, A., Easterly, A., Kurlat, S., and Wacziarg, R. (2003) "Fractionalization," *Journal of Economic Growth*, Vol. 8, 155-194.

[5] Austen, R.A. (1987) **African Economic History: Internal Development and External Dependency**, James Currey.

[6] Azevado, M.J. (1980) "Epidemic Disease among the Sara of Southern Chad, 1890-1940," in Hartwig, G.W. and Patterson, K.D., **Disease in African History**, Duke University Press, 119-152.

[7] Bhattacharyya , S. (2009) "Root Causes of African Underdevelopment," *Journal of African Economies*, Vol. 18, 745-780.

[8] Biraben, J-N. (1979) "Essai sur L'évolution du Nombre des Hommes," *Population*, Vol. 34, 13-25.

[9] Bockstette, V., Chanda, A., and Putterman, L. (2002) "States and Markets: The Advantage of an Early Start," *Journal of Economic Growth*, Vol. 7, 347-369.

[10] Bloom, D.E. and Sachs, J. (1998) "Geography, Demography, and Economic Growth in Africa," *Brookings Papers on Economic Activity*, Economic Studies Program, Vol. 29, 207-296.

[11] Cashdan, E. (2001) "Ethnic Diversity and its Environmental Determinants: Effects of Climate, Pathogens, and Habitat Diversity," *American Anthropologist*, Vol. 103, 968-991.

[12] Chandler, T. (1987) **Four Thousand Years of Urban Growth: An Historical Census**, St. Davis University Press.

[13] Collier, P. and Gunning, J.W. (1999) "Explaining African Economic Performance," *Journal of Economic Literature*, Vol. 37, 64-111.

[14] Connah, G. (2001) **African Civilizations: An Archaeological Perspective**, Cambridge University Press.

[15] Curtin, P., Feierman, S., Thompson, L., and Vansina, J. (1995) **African History: From Earliest Times to Independence**, Longman: New York.

[16] Davidson, B. (1959) **The Lost Cities of Africa**, Little, Brown and Company.

[17] Diamond, J. (1997) **Guns, Germs and Steel: The Fates of Human Societies**, Vintage.

[18] Easterly, W. and Levine, R. (1997) "Africa's Growth Tragedy: Policies and Ethnic Divisions," *Quarterly Journal of Economics*, Vol. 112, 1203-1250.

[19] Englebert, P. (2000) **State Legitimacy and Development in Africa**, Lynne Rienner Publishers.

[20] Fincher, C.L. and Thornhill, R. (2008) "A Parasite-driven Wedge: Infectious Diseases may Explain Language and other Biodiversity," *Oikos*, Vol. 117, 1289 1297.

[21] Fortes, M. and Evans-Pritchard, E.E. (1940) **African Political Systems**, Oxford University Press.

[22] Gennaioli, N. and Rainer, I. (2007) "The Modern Impact of Precolonial Centralization in Africa," *Journal of Economic Growth*, Vol. 12, 185-234.

[23] Herbst, J. (2000) **States and Power in Africa: Comparative Lessons in Authority and Control**, Princeton University Press.

[24] Hopkins, A.G. (1973) **An Economic History of West Africa**, Columbia University Press.

[25] Kamarck, A. (1967) **The Economics of African Development**, Praeger Publishers.

[26] Hull, R.W. (1976) **African Cities and Towns before the European Conquest**, W.W. Norton.

[27] Low, B.S. (1990) "Marriage Systems and Pathogen Stress in Human Societies," *American Zoologist*, Vol. 30, 325-339.

[28] Maddison, A. (2001) **The World Economy: A Millennial Perspective**, OECD Development Centre.

[29] Mann, C.C. (2005) **1491: New Revelations of the Americas Before Columbus**, Alfred A. Knopf.

[30] McEvedy, C. and Jones, R. (1985) **Atlas of World Population History**, Penguin Books.

[31] McNeill, W.H. (1976) **Plagues and People**, Anchor Books.

[32] Michalopolous, S. (2008) "The Origins of Ethnolinguistic Diversity: Theory and Evidence," Working Paper.

[33] Miguel, E. and Gugerty, M.K. (2005) "Ethnic Diversity, Social Sanctions, and Public Goods in Kenya," *Journal of Public Economics*, Vol. 89., 2325-2368.

[34] Modelski, G. (2003) **World Cities:−3000 to 2000**, Faros 2000.

[35] Murdock, G.P. and White, D.R. (1969) "Standard Cross-Cultural Sample," *Ethnology*, Vol. 9, 329-369. Updates are available online at http://repositories.cdlib.org/imbs/socdyn/wp/Standard_Cross-Cultural_Sample

[36] Nunn, N. (2008) "The Long-Term Effects of Africa's Slave Trades," *Quarterly Journal of Economics*, Vol. 123, 139-176.

[37] Patterson, O. (1982) **Slavery and Social Death: A Comparative Study**, Harvard University Press.

[38] Pryor, F.L. (2005) **Economic Systems of Foraging, Agricultural, and Industrial Societies**, Cambridge University Press.

[39] Smith, B.D. (1995) **The Emergence of Agriculture**, Scientific American Library.

[40] Spolaore, E. and Wacziarg, R. (2009) "The Diffusion of Development," *Quarterly Journal of Economics*, Vol. 124, 469-529.

[41] Vengroff, R. (1976) "Population Density and State Formation in Africa," *African Studies Review*, Vol. 19, 67-74.

[42] Weber, E. (1976) **Peasants into Frenchmen: The Modernization of Rural France, 1870-1914**, Stanford University Press.

[43] Wolfe, N.D., Dunavan, C.P., and Diamond, J. (2007) "Origins of Major Human Infectious Diseases," *Nature*, Vol. 447, 279-283.

# 6 Appendix A: Alternative *proxies*

Here I report two *proxies* for economic and political sophistication in pre-modern times. Following Acemoglu et al. ([2], Section 2), I first examine urbanization.

Table A1 reports the number of cities with populations over 20 and 40 thousand inhabitants in sub-Saharan Africa and South and Central America from Chandler [12]. (The inventory in Chandler [12], according to Connah [14], provided accurate patterns of city formation in Africa.) Table A1 reports different time periods and divides sub-Saharan Africa into three sub-regions. The cities in regions with high Arab influence are coded as Muslims while the Middle Nile and Ethiopia are regions influenced by ancient Egypt. The rest of sub-Saharan Africa can be consider as indigenous city formation. The number of cities in areas without Arab influence is always smaller than in tropical America.[15] After 1200, the Arab influence in Africa increased substantially. In 1500 there were 13 cities in regions with Arab influence. These cities, however, were heavily concentrated in West Africa. In 1500, the number of cities with more than 40 thousand inhabitants was the same in both regions.

Table A1. Cities in Africa and the New World.

| Year | North Africa | Sub-Saharan Africa: Muslims | Sub-Saharan Africa: Middle Nile and Ethiopia | Sub-Saharan Africa: Rest (indigenous) | Sub-Saharan Africa: Total | North America | South and Central America |
|---|---|---|---|---|---|---|---|
| | A. Number of cities with populations over 20,000 inhabitants | | | | | | |
| 800 | 10 | 0 | 2 | 3 | 5 | 0 | 10 |
| 1000 | 13 | 0 | 1 | 4 | 5 | 0 | 9 |
| 1200 | 18 | 6 | 2 | 4 | 12 | 0 | 10 |
| 1300 | 18 | 8 | 2 | 5 | 15 | 0 | 11 |
| 1400 | 18 | 8 | 2 | 9 | 19 | 0 | 18 |
| 1500 | 19 | 13 | 3 | 8 | 24 | 1 | 16 |
| | B. Number of cities with populations over 40,000 inhabitants | | | | | | |
| 800 | 4 | 0 | 0 | 1 | 1 | 0 | 2 |
| 1500 | 7 | 4 | 0 | 2 | 6 | 0 | 6 |

Note.– Data from Chandler ([12], 39-57). The indigenous cities in sub-Saharan Africa cover mostly Ghana, Zimbabwe and the Bantus. The middle Nile corresponds to Dongola (modern Sudan) and Kaffa. North Africa includes cities in the Mediterranean (i.e., Arabian, Egypt, Spanish Africa, and Aloa) and the Maghreb.

Population size and density are alternative *proxies* commonly used to study pre-modern economies. Table A2 reports data from Biraben [8] and McEvedy and Jones [30]. These estimates differ on some dimensions but in both, sub-Saharan Africa has larger populations in 1500 than

[15] Ancient Egypt appently did not have a strong influence on sub-Saharan Africa although cities appeared in the middle Nile earlier than in other areas. Meroë is the best known case. In 430 B.C., Meroë had about 20,000 inhabitants (Chandler [12], 461). Gold, ivory, slaves, and other mineral, animal and vegetable products were traded with Eurasia through the Nubian corridor that connected tropical Africa with Egypt. It has been suggested that this corridor was a cultural *cul-de-sac*, Connah ([14], 19).

the Americas, and Africa has the fastest population growth in the world in the years between 400 B.C. (or A.D.) and 1500 (or 1000).

Table A1 suggests that cities in sub-Saharan Africa had more recent and external origins, and that they were scarcer and less densely populated than those in tropical America or in North Africa. Given the differences in population sizes, urbanization rates in 1500 were also lower in Africa. Multiply the number of cities at each size by the cut-off size to obtain an estimate of the size of urban populations. For South and Central America this number in 1500 is $10 \times 20,000 + 6 \times 40,000 = 440,000$. If the Muslim cities are counted as part of Africa, the same estimate for sub-Saharan Africa is $600,000$. If Muslim cities are excluded, the size of the urban population is $260,000$.

Table A2. Estimated population in Africa and the Americas.

| | | Biraben [8] | | | | McEvedy and Jones [30] | | |
|---|---|---|---|---|---|---|---|---|
| Region | Area | 400 B.C. | A.D. | 1000 | 1500 | A.D. | 1000 | 1500 |
| Africa | | | | | | | | |
| North | 2 | 10 | 14 | 9 | 9 | 8 | 11 | 8 |
| Sub-Saharan | 25 | 7 | 12 | 30 | 78 | 8 | 22 | 38 |
| The Americas | | | | | | | | |
| North | 20 | 1 | 2 | 2 | 3 | 0.4 | 0.7 | 1.3 |
| South and Central | 20 | 7 | 10 | 16 | 39 | 4 | 8 | 13 |
| World population | | 153 | 252 | 253 | 461 | 170 | 265 | 425 |

Note.– Population in millions. Area (mill. km$^2$) from McEvedy and Jones [30]. North Africa includes the Maghreb, Libya and Egypt. The area in North Africa does not include the Sahara. North America includes the US, Canada, and the Caribbean.

Using population size from Biraben [8] gives urbanization rates for tropical America of 1.12 percent and for Africa it gives 0.76 percent (or 0.3 percent if Arab cities are excluded). Rescaling the previous rates using the average urbanization rate for the 22 countries in South and Central America in 1500 listed in the Appendix 3 of Acemoglu et al. [2] yields an urbanization rate for Africa of 3.96 percent (or 1.71 percent if Arab cities are excluded). Using population size from McEvedy and Jones [30], the estimates are 2.70 and 1.17 percent.

Based on the relationship between urbanization and GDP in Acemoglu et al. ([2], Table 2), the smallest predicted relative income difference between the Americas and Africa in 1500 is $\exp\{0.038 \times 5.80/3.96\} = 1.06$. Using 550 as the income in Latin America implies that in Africa is $550/1.06 = 525$. The highest predicted difference generates an income of 455.

Further, Acemoglu et al. ([2], Table 3 (1)) provides a point estimate of urbanization rates in 1500 on current GDP per capita for former European colonies of $-0.078$ (standard error 0.026). Using this estimate to determine current income rankings gives contradictory predictions for Africa.[16] Because urbanization rates in Africa were smaller than in the Americas, predictions based on the previous relationship suggest Africa should have an income per capita between 20 to 40 percent higher than in Latin America in 2000, \$7584 to \$9950.

[16] It is important to stress that the estimates in Acemoglu et al. [2] do not include sub-Saharan Africa so the prediction for urbanization is *out of sample*.

It is relevant to explain why the *reversal of fortune* in Acemoglu et al. [2] does not account for Africa's poor performance. Acemoglu et al. [2] provide an important account for why the colonies of North America and Australia are today an order of magnitude richer than Latin America and Africa. These colonies were less urbanized and hence provided fewer incentives for extractive institutions. Their analysis may not be as informative as to why within the tropics Africa is lagging behind Latin America. Africa and the Latin America were much more similar in the past than any of these regions and North America or Australia. As Acemoglu et al. ([2], 1238) noted, urbanization in sub-Saharan Africa "was at a higher level than in North America or Australia." Thus, in order to account for the low incomes seen in Africa today using a reversal of fortune argument, one would have to argue that societies in Africa were far more urbanized than societies in Latin America. If the urbanization rates in Africa were in between the high urbanization rates of South and Central America and the low urbanization rates of North America, current incomes in Africa should also lie in the middle of the world income distribution.[17]

[17] It is also important to notice that the difference in population density contradicts Herbst [23], who argued that the lack of state consolidation in pre-colonial Africa was in part a consequence of a relatively low population density, which made state control more difficult and competition for space less attractive. Implicit in Herbst [23] is a comparison between state formation in Africa and Europe. The comparison between pre-colonial Africa and the Americas suggests a weaker link between state formation and population densities. The Americas had smaller population densities than Africa but the independent development of pre-Columbian states in the Americas resembles to an astonishing degree those patterns seen in the earliest states of Eurasia.

# 7 Appendix B: Robustness checks [NOT NEEDED FOR PUBLICATION]

**Results complementing Table 6**. Table A1 presents robustness checks for Table 6, panel A. These results show that the negative OLS correlation between ethnic diversity and large or impressive structures is robust to additional controls and sub-samples.

Table A1. Robustness checks for Table 6, panel A. OLS estimates.

| | Dependent variable: Large or impressive structures | | | | |
|---|---|---|---|---|---|
| | (2) | (3) | (4) | (5) | (6) |
| | A. An alternative measure of ethnic diversity | | | | |
| Ethnic groups | −0.039*** | −0.030** | −0.028*** | −0.026*** | −0.021** |
| within 250 m. | (0.009) | (0.010) | (0.009) | (0.009) | (0.010) |
| N. Obs. | 127 | 123 | 123 | 123 | 123 |
| | B. Using religious, military, or industrial structures only | | | | |
| Ethnic groups | −0.003*** | −0.003** | −0.003** | −0.002** | −0.002 |
| within 500 m. | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| N. Obs. | 127 | 123 | 123 | 123 | 123 |
| | C. Controlling for density instead of population size | | | | |
| Ethnic groups | −0.015*** | −0.012*** | −0.012*** | −0.011*** | −0.009** |
| within 500 m. | (0.003) | (0.004) | (0.003) | (0.003) | (0.004) |
| N. Obs. | 126 | 122 | 122 | 122 | 122 |
| | D. Societies in sub-Saharan Africa and tropical America | | | | |
| Ethnic groups | −0.013*** | −0.015*** | −0.016*** | −0.015*** | −0.013** |
| within 500 m. | (0.003) | (0.004) | (0.004) | (0.004) | (0.005) |
| N. Obs. | 52 | 50 | 50 | 50 | 50 |
| | E. Societies in sub-Saharan Africa and Eurasia | | | | |
| Ethnic groups | −0.018*** | −0.015*** | −0.014*** | −0.014*** | −0.012*** |
| within 500 m. | (0.003) | (0.004) | (0.004) | (0.004) | (0.004) |
| N. Obs. | 101 | 97 | 97 | 97 | 97 |
| | F. Excluding societies in the Sahel | | | | |
| Ethnic groups | −0.013*** | −0.012*** | −0.013*** | −0.012*** | −0.009** |
| within 500 m. | (0.003) | (0.004) | (0.004) | (0.004) | (0.004) |
| N. Obs. | 122 | 118 | 118 | 118 | 118 |
| | G. Societies in Africa (North and sub-Saharan) | | | | |
| Ethnic groups | −0.016*** | −0.011 | −0.012 | −0.012 | −0.012 |
| within 500 m. | (0.005) | (0.009) | (0.008) | (0.008) | (0.009) |
| N. Obs. | 35 | 33 | 33 | 33 | 33 |
| | H. Excluding societies in sub-Saharan Africa | | | | |
| Ethnic groups | −0.003 | −0.001 | 0.0006 | 0.002 | 0.002 |
| within 500 m. | (0.008) | (0.009) | (0.008) | (0.009) | (0.009) |
| N. Obs. | 101 | 99 | 99 | 99 | 99 |

Note.– In parentheses are robust standard errors. ***, ** and * denote statistical significance at the 1, 5, and 10 percent level. The specifications follow those of Table 5.

Table A.2 presents similar robustness checks for Table 6, panel B. These results rely on an IV strategy which uses pathogen stress as instrument for ethnic diversity.

Table A2. Robustness checks for Table 6, panel B. IV estimates (second-stage).

| | Dependent variable: Large or impressive structures | | | | |
|---|---|---|---|---|---|
| | (2) | (3) | (4) | (5) | (6) |
| | A. An alternative measure of ethnic diversity | | | | |
| Ethnic groups | −0.080*** | −0.097** | −0.114*** | −0.113*** | −0.133** |
| within 250 m. | (0.026) | (0.048) | (0.041) | (0.042) | (0.063) |
| N. Obs. | 127 | 123 | 123 | 123 | 123 |
| | B. Using religious, military, or industrial structures only | | | | |
| Ethnic groups | −0.005** | −0.006 | −0.009** | −0.009** | −0.010** |
| within 500 m. | (0.002) | (0.004) | (0.003) | (0.003) | (0.005) |
| N. Obs. | 127 | 123 | 123 | 123 | 123 |
| | C. Controlling for density instead of population size | | | | |
| Ethnic groups | −0.031*** | −0.034** | −0.039*** | −0.038*** | −0.042** |
| within 500 m. | (0.008) | (0.014) | (0.013) | (0.013) | (0.019) |
| N. Obs. | 126 | 122 | 122 | 122 | 122 |
| | D. Societies in sub-Saharan Africa and tropical America | | | | |
| Ethnic groups | −0.019*** | −0.019* | −0.021* | −0.021* | −0.022 |
| within 500 m. | (0.006) | (0.011) | (0.011) | (0.011) | (0.020) |
| N. Obs. | 52 | 50 | 50 | 50 | 50 |
| | E. Societies in sub-Saharan Africa and Eurasia | | | | |
| Ethnic groups | −0.026*** | −0.034** | −0.038*** | −0.038*** | −0.040** |
| within 500 m. | (0.007) | (0.016) | (0.012) | (0.012) | (0.016) |
| N. Obs. | 101 | 97 | 97 | 97 | 97 |
| | F. Excluding societies in the Sahel | | | | |
| Ethnic groups | −0.020*** | −0.030** | −0.032*** | −0.032*** | −0.035** |
| within 500 m. | (0.006) | (0.010) | (0.010) | (0.010) | (0.014) |
| N. Obs. | 122 | 118 | 118 | 118 | 118 |
| | G. Societies in Africa (North and sub-Saharan) | | | | |
| Ethnic groups | −0.044** | −0.126 | −0.115 | −0.116 | −1.075* |
| within 500 m. | (0.017) | (0.087) | (0.070) | (0.071) | (0.058) |
| N. Obs. | 35 | 33 | 33 | 33 | 33 |
| | H. Excluding societies in sub-Saharan Africa | | | | |
| Ethnic groups | −0.033 | 0.051 | −0.832 | −0.573 | −0.483 |
| within 500 m. | (0.878) | (0.227) | (8.947) | (3.915) | (2.814) |
| N. Obs. | 101 | 99 | 99 | 99 | 99 |

Note.– In parentheses are robust standard errors. ***, ** and * denote statistical significance at the 1, 5, and 10 percent level. The specifications follow those of Table 5.

**Results complementing Table 8**. Table A3 presents robustness checks for Table 8, panel A. The table presents OLS correlations between pathogen stress and current income per capita for alternative specifications, controls, and sub-samples.

Table A3. Robustness checks for Table 8, panel A. Pathogen stress and GDP per capita.

| | Dependent variable: log-GDP per capita | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| | A. Using the presence of pathogens only | | | | | | |
| Pathogen presence | −0.409*** | −0.305*** | −0.364*** | −0.267*** | −0.273*** | −0.258*** | −0.254*** |
| | (0.04) | (0.05) | (0.06) | (0.06) | (0.06) | (0.06) | (0.07) |
| N. Obs. | 67 | 66 | 67 | 66 | 65 | 66 | 66 |
| | B. Adding food surplus and tech. sophistication as controls | | | | | | |
| Pathogen stress | −0.172*** | −0.155*** | −0.152*** | −0.128*** | −0.131*** | −0.101** | −0.08* |
| | (0.03) | (0.03) | (0.03) | (0.03) | (0.02) | (0.04) | (0.04) |
| N. Obs. | 67 | 66 | 67 | 66 | 65 | 66 | 66 |
| | C. Including non-colonies | | | | | | |
| Pathogen stress | −0.234*** | −0.233*** | −0.207*** | −0.207*** | −0.190*** | −0.199*** | −0.121** |
| | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) | (0.04) |
| N. Obs. | 89 | 88 | 89 | 88 | 87 | 87 | 88 |
| | D. Excluding sub-Saharan Africa | | | | | | |
| Pathogen stress | −0.197*** | −0.101** | −0.180*** | −0.082** | −0.091** | −0.031 | −0.052 |
| | (0.03) | (0.03) | (0.04) | (0.04) | (0.04) | (0.04) | (0.06) |
| N. Obs. | 40 | 39 | 40 | 39 | 39 | 39 | 39 |
| | E. Only Africa (sub-Saharan and North) | | | | | | |
| Pathogen stress | −0.112** | −0.111** | −0.134*** | −0.139*** | −0.141*** | −0.169** | 0.101 |
| | (0.05) | (0.05) | (0.04) | (0.04) | (0.05) | (0.06) | (0.10) |
| N. Obs. | 32 | 32 | 32 | 32 | 31 | 32 | 32 |
| | F. Sub-Saharan Africa and South and Central America | | | | | | |
| Pathogen stress | −0.191*** | −0.166*** | −0.116*** | −0.101** | −0.100** | −0.098* | −0.009 |
| | (0.03) | (0.03) | (0.03) | (0.04) | (0.03) | (0.05) | (0.06) |
| N. Obs. | 51 | 51 | 51 | 51 | 50 | 51 | 51 |

Note.– In parentheses are robust standard errors. ***, ** and * denote statistical significance at the 1, 5, and 10 percent level. Each specification corresponds to those of Table 8.

Table A4 presents robustness checks for Table 8, panel B, which pertain to the OLS correlations between pre-industrial ethnic diversity and current income per capita.

Table A4. Robustness checks for Table 8, panel B. Ethnic diversity and GDP per capita.

| | Dependent variable: log-GDP per capita | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| | A. Adding food surplus and tech. sophistication as controls | | | | | | |
| Ethnic groups in 500 m. | −0.014*** | −0.014*** | −0.011*** | −0.012*** | −0.012** | −0.008* | −0.005 |
| | (0.003) | (0.003) | (0.04) | (0.004) | (0.005) | (0.004) | (0.004) |
| N. Obs. | 67 | 66 | 67 | 66 | 65 | 66 | 66 |
| | B. Using an alternative measure of ethnic diversity | | | | | | |
| Ethnic groups in 250 m. | −0.056*** | −0.038*** | −0.041*** | −0.026** | −0.027** | −0.016 | −0.010 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| N. Obs. | 67 | 66 | 67 | 66 | 65 | 66 | 66 |
| | C. Including non-colonies | | | | | | |
| Ethnic groups in 500 m. | −0.023*** | −0.023*** | −0.019*** | −0.019*** | −0.020*** | −0.016*** | −0.008* |
| | (0.003) | (0.003) | (0.004) | (0.004) | (0.005) | (0.005) | (0.005) |
| N. Obs. | 87 | 86 | 87 | 86 | 85 | 85 | 86 |
| | D. Excluding sub-Saharan Africa | | | | | | |
| Ethnic groups in 500 m. | −0.015 | −0.005 | −0.019 | −0.010 | −0.015 | −0.016 | 0.007 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| N. Obs. | 40 | 39 | 40 | 39 | 39 | 39 | 39 |
| | E. Only Africa (sub-Saharan and North) | | | | | | |
| Ethnic groups in 500 m. | −0.009** | −0.008** | −0.009* | −0.008* | −0.008 | −0.006 | 0.015 |
| | (0.004) | (0.004) | (0.005) | (0.004) | (0.007) | (0.005) | (0.007) |
| N. Obs. | 32 | 32 | 32 | 32 | 31 | 32 | 32 |
| | F. Sub-Saharan Africa and South and Central America | | | | | | |
| Ethnic groups in 500 m. | −0.017*** | −0.014*** | −0.008** | −0.007 | −0.006 | −0.005 | 0.002 |
| | (0.003) | (0.003) | (0.004) | (0.005) | (0.005) | (0.005) | (0.005) |
| N. Obs. | 51 | 51 | 51 | 51 | 50 | 51 | 51 |

Note.– In parentheses are robust standard errors. ***, ** and * denote statistical significance at the 1, 5, and 10 percent level. Each specification corresponds to those of Table 8.

Table A5 presents robustness checks for Table 8, panel C, which pertain to the IV estimates of pre-industrial ethnic diversity and current income per capita.

Table A5. Robustness checks for Table 8, panel C. Diversity and income (second-stage).

| | Dependent variable: log-GDP per capita | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| | A. Using the presence of pathogens only | | | | | | |
| Ethnic groups | −0.045*** | −0.033*** | −0.048*** | −0.034*** | −0.035*** | −0.030** | −0.038 |
| in 500 m. | (0.012) | (0.009) | (0.016) | (0.012) | (0.012) | (0.013) | (0.024) |
| N. Obs. | 67 | 66 | 67 | 66 | 65 | 66 | 66 |
| | B. Adding food surplus and tech. sophistication as controls | | | | | | |
| Ethnic groups | −0.033*** | −0.029*** | −0.036*** | −0.029*** | −0.030*** | −0.023** | −0.022 |
| in 500 m. | (0.009) | (0.008) | (0.013) | (0.009) | (0.010) | (0.012) | (0.014) |
| N. Obs. | 67 | 66 | 67 | 66 | 65 | 66 | 66 |
| | C. Using an alternative measure of ethnic diversity | | | | | | |
| Ethnic groups | −0.150*** | −0.111*** | −0.141*** | −0.097*** | −0.107** | −0.071** | −0.083* |
| in 250 m. | (0.040) | (0.032) | (0.047) | (0.036) | (0.040) | (0.033) | (0.047) |
| N. Obs. | 67 | 66 | 67 | 66 | 65 | 66 | 66 |
| | D. Including non-colonies | | | | | | |
| Ethnic groups | −0.047*** | −0.046*** | −0.047*** | −0.046*** | −0.038*** | −0.042*** | −0.035* |
| in 500 m. | (0.009) | (0.008) | (0.011) | (0.010) | (0.007) | (0.010) | (0.018) |
| N. Obs. | 87 | 86 | 87 | 86 | 85 | 85 | 86 |
| | E. Excluding sub-Saharan Africa | | | | | | |
| Ethnic groups | −0.290 | −0.225 | −0.205 | −0.116 | −0.118 | −0.035 | −0.130 |
| in 500 m. | (0.403) | (0.431) | (0.223) | (0.152) | (0.115) | (0.060) | (0.411) |
| N. Obs. | 40 | 39 | 40 | 39 | 39 | 39 | 39 |
| | F. Only Africa (sub-Saharan and North) | | | | | | |
| Ethnic groups | −0.018** | −0.017** | −0.022** | −0.022** | −0.037** | −0.024** | −0.141 |
| in 500 m. | (0.008) | (0.008) | (0.008) | (0.008) | (0.013) | (0.011) | (0.686) |
| N. Obs. | 32 | 32 | 32 | 32 | 31 | 32 | 32 |
| | F. Sub-Saharan Africa and South and Central America | | | | | | |
| Ethnic groups | −0.030*** | −0.028*** | −0.024*** | −0.022** | −0.024** | −0.018** | −0.007 |
| in 500 m. | (0.006) | (0.007) | (0.009) | (0.009) | (0.010) | (0.009) | (0.048) |
| N. Obs. | 51 | 51 | 51 | 51 | 50 | 51 | 51 |

Note.– In parentheses are robust standard errors. ***, ** and * denote statistical significance at the 1, 5, and 10 percent level. Each specification corresponds to those of Table 8.

**Results complementing Table 9**. Table A6 presents robustness checks for Table 9.

Table A6. Robustness checks for Table 9, panel I.A. Pathogen stress and ELF.

| | Dependent variable: Ethnolinguistic fractionalization | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| | A. Using the presence of pathogens only | | | | | | |
| Pathogen stress | 0.061*** | 0.063*** | 0.057*** | 0.061*** | 0.065*** | 0.073** | 0.029 |
| | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) | (0.03) |
| N. Obs. | 66 | 65 | 66 | 65 | 65 | 65 | 65 |
| | B. Adding food surplus and tech. sophistication as controls | | | | | | |
| Pathogen stress | 0.033*** | 0.035*** | 0.026** | 0.028** | 0.030** | 0.031* | −0.012 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| N. Obs. | 71 | 68 | 71 | 68 | 68 | 68 | 68 |
| | C. Including non-colonies | | | | | | |
| Pathogen stress | 0.043*** | 0.045*** | 0.040*** | 0.043*** | 0.048*** | 0.044*** | −0.007 |
| | (0.007) | (0.007) | (0.008) | (0.008) | (0.009) | (0.008) | (0.015) |
| N. Obs. | 87 | 84 | 87 | 84 | 84 | 84 | 84 |
| | D. Excluding sub-Saharan Africa | | | | | | |
| Pathogen stress | 0.001 | −0.006 | 0.002 | −0.006 | −0.032 | −0.006 | −0.026 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.02) | (0.03) |
| N. Obs. | 44 | 41 | 44 | 41 | 41 | 41 | 41 |
| | E. Only Africa (sub-Saharan and North) | | | | | | |
| Pathogen stress | 0.019 | 0.030** | 0.020 | 0.035** | 0.029 | 0.059** | −0.044 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.02) | (0.02) | (0.03) |
| N. Obs. | 34 | 33 | 34 | 33 | 33 | 33 | 33 |
| | F. Sub-Saharan Africa and South and Central America | | | | | | |
| Pathogen stress | 0.045*** | 0.047*** | 0.021 | 0.027* | 0.026* | 0.026 | −0.008 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.02) | (0.02) | (0.02) |
| N. Obs. | 53 | 52 | 52 | 52 | 52 | 52 | 52 |

Note.– In parentheses are robust standard errors. ***, ** and * denote statistical significance at the 1, 5, and 10 percent level. Each specification corresponds to those of Table 8.

Table A7 presents robustness checks for Table 9, panel C. IV estimates are provided for cases when the first-stage results in Table A6 are significant.

Table A7. Robustness checks for Table 9, panel II.B. IV of pathogen stress and ELF.

| | Dependent variable: log-GDP per capita | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| | A. Using the presence of pathogens only | | | | | | |
| ELF | −6.556*** | −4.716*** | −6.206*** | −4.156** | −4.185*** | −3.402** | −7.544 |
| | (1.80) | (1.55) | (1.85) | (1.56) | (1.42) | (1.68) | (8.48) |
| N. Obs. | 66 | 65 | 66 | 65 | 65 | 65 | 65 |
| | B. Adding food surplus and tech. sophistication as controls | | | | | | |
| ELF | −5.054*** | −4.382*** | −5.589** | −4.328** | −4.240*** | −3.220* | 9.774 |
| | (1.73) | (1.42) | (2.32) | (1.74) | (1.55) | (1.91) | (27.87) |
| N. Obs. | 66 | 65 | 66 | 65 | 65 | 65 | 65 |
| | C. Including non-colonies | | | | | | |
| ELF | −5.523*** | −5.243*** | −5.136*** | −4.772*** | −4.185*** | −4.495*** | - |
| | (0.98) | (0.87) | (1.08) | (0.90) | (0.80) | (0.92) | - |
| N. Obs. | 82 | 81 | 82 | 81 | 81 | 81 | - |

Note.– In parentheses are robust standard errors. ***, ** and * denote statistical significance at the 1, 5, and 10 percent level. Each specification corresponds to those of Table 8. ELF denotes ethnolinguistic fractionalization. Specifications C–E in Table A6 are not included because the first-stage results are not valid for an instrumental varible estimate. Simialry, - denotes invalid first-stage estimates in specification B (7).

I also investigated (but do not report here) systematic differences between hunter-gatherers in Africa and South and Central America to address concerns with prejudice Westerners may have about Africa and with sample selection. Since most of the population movements in the long pre-agricultural period have been out of Africa, the sample of societies in Africa may be a negatively selected sample of hunter-gatherers whereas the hunter-gatherers that migrated into the Americas had positive attributes that later conduced to faster agricultural developments. Further, because most of the societies in Africa were described at the peak of colonization, there may be prejudice against African societies. There are no predictable differences between the hunter-gatherers in Africa and those in South and Central America. This suggests that prejudice and selection may not be crucial concerns.[18]

[18] I also excluded Haitians (#160) and Saramaca (#165) from the Americas since they may have some African influence but their exclusion does not alter any result hence I do not report these estimates. Perhaps it is important to notice that I have excluded Madagascar from the sub-Saharan sample. Despite the proximity to East Africa, the first human settlements of Madagascar came from Asia.

# 8 Appendix C: Country codes for the SCCS [NOT NEEDED FOR PUBLICATION]

1 = ZAF,NAM,BWA, 2 = NAM,BWA,AGO, 3 = MOZ,ZAF, 4 = ZMB,NAM,AGO,BWA, 5 = AGO,6 = CGO,AGO, 7 = ZMB, 8 = TZA,MWI, 9 = TZA, 10 = TZA, 11 = KEN, 12 = UGA, 13 = CGO, 14 = CGO, 15 = CMR, 16 = NGA,CMR, 17 = NGA, 18 = BEN,NGA, 19 = GHA, 20 = SLE, 21 = SEN,GMB,MRT, 22 = MLI,BFA,SEN, 23 = GHA, 24 = MLI, 25 = NER,NGA,CMR,26 = NGA,NER,SDN,CMR,GHA,CIV,TCD, 27 = TCD,CMR, 28 = ZAR,SDN,CAF, 29 = SDN, 30 = SDN, 31 = SDN, 32 = ETH, 33 = ETH, 34 = KEN,TZA, 35 = ETH, 36 = SOM,ETH,YEM,KEN,DJI, 37 = ETH, 38 = ERI, 39 = SDN,EGY, 40 = TCD,LBY,NER,NGA, 41 = NER,MLI,DZA,BFA,LBY, 42 = MAR, 43 = EGY, 44 = ISR, 45 = IRQ, 46 = SAU, 47 = TUR, 48 = ALB, 49 = ITA, 50 = ESP,FRA, 51 = IRL, 52 = NOR,SWE,FIN, 53 = RUS, 54 = RUS, 55 = TUR,RUS, 56 = ARM,RUS, 57 = TUR,IRN,IRQ, SYR, 58 = IRN, 59 = PAK,IND, 60 = IND, 61 = IND, 62 = IND,BGD, 63 = IND,NPL, 64 = PAK, 65 = KAZ,UZB,CHN,MNG, 66 = MNG, 67 = CHN,VNM,THA, 68 = IND,NPL,BTN, 69 = IND,BGD, 70 = IND,BUR, 71 = BUR, 72 = LAO, 73 = VNM, 74 = VNM, 75 = KHM, 76 = THA, 77 = BUR,MYS,SGP,THA, 78 = IND, 79 = IND, 80 = LKA, 81 = MDG, 82 = MYS, 83 = IDN, 84 = IDN, 85 = MYS, 86 = MYS,PHL, 87 = IDN, 88 = IDN, 89 = IDN, 90 = AUS, 91 = AUS, 92 = PNG, 93 = IDN, 94 = IDN, 95 = PNG, 96 = PNG, 97 = PNG, 98 = PNG, 99 = PNG, 100 = SLB, 101 = VUT, 102 = FJI,NZL 103 = PYF, 104 = NZL,AUS, 105 = PYF, 106 = WSM, 107 = KIR, 108 = MHL, 109 = FSM, 110 = FSM, 111 = PLW, 112 = PHL, 113 = CHN, 114 = CHN, 115 = CHN, 116 = KOR, 117 = JPN, 118 = JPN,RUS, 119 = RUS, 120 = RUS, 121 = RUS, 122 = USA, 123 = USA, 124 = CAN, 125 = CAN, 126 = CAN,USA, 127 = CAN, 128 = CAN, 129 = CAN, 130 = USA, 131 = CAN,USA, 132 = CAN, 133 = USA, 134 = USA, 135 = USA, 136 = USA, 137 = USA, 138 = USA, 139 = CAN,USA, 140 = USA, 141 = USA, 142 = USA, 143 = USA, 144 = CAN,USA, 145 = USA, 146 = USA, 147 = USA, 148 = USA,MEX, 149 = USA, 150 = USA, 151 = USA,MEX, 152 = MEX, 153 = MEX, 154 = MEX, 155 = GTM, 156 = HND,NIC, 157 = CRI, 158 = PAN,COL, 159 = COL,VEN, 160 = HTI, 161 = DMA, 162 = VEN,GUY, 163 = VEN,BRA, 164 = GUY, 165 = SUR, 166 = BRA, 167 = BRA,COL, 168 = ECU,COL, 169 = ECU,PER, 170 = PER,BRA, 171 = PER, 172 = BOL,PER,CHI, 173 = BOL, 174 = BRA, 175 = BRA, 176 = BRA, 177 = BRA, 178 = BRA, 179 = BRA, 180 = BRA, 181 = PRY,URY,BRA, 182 = PRY, 183 = ARG, 184 = CHL, 185 = ARG, 186 = ARG,CHL.