Title
The Current Economic Impact of Variation in Early Disease Environments

Author
Ward, Emily

Publication Date
2016-10-24

The Current Economic Impact of Variation in Early Disease Environments

Emily Ward

Advisor: Javier Birchenall

INTRODUCTION

This paper studies the possible correlation between early European mortality rates in colonies and the early institutions that Europeans set up in these same places. It is essentially a replication of Acemoglu, Johnson and Robinson's (henceforth known as AJR) "The Colonial Origins of Comparative Development: An Empirical Investigation" (2000), which used settler mortality rates as an instrument for institutions. AJR theorized that where European colonists faced harsher disease environments upon first arrival they were more likely to leave behind so-called bad—or extractive—institutions, which carried over into the present. For example, they reasoned that there are, in part, such large income disparities between colonies that became the likes of the United States, or Australia, and the colonies of almost every current African nation because early settlers faced a much higher risk of mortality in the later areas.

My study differs from AJR's by using new, comprehensive data in order to calculate more accurate and representative mortality rates for the European settlers. As census records are scarce for actual European settlers, AJR used data from European soldiers and bishops in order to calculate mortality rates hypothetically representative of settlers. My data comes from two separate sources, over six thousand missionaries and over twenty thousand Catholic bishops. This is an attempt to reconcile concerns that economists had with AJR's original data, the main one being that it was lack of observations for each colony studied. Though their main focus was African colonies, AJR did not have actual records for about half of the colonies that they calculated mortality rates for, forcing them to speculatively fill in the holes with what data they did have. This led to the worry that there were sometimes contradictory or unjustified reasons behind using mortality rates from one colony as a representation of rates from an entirely different colony (Albouy, 2012). This practice could have easily skewed the calculated effects of mortality rates in favor of AJR's arguments. The data I have collected is much more varied and contains actual observations from missionaries or bishops in almost every location, so issues such as these can be avoided. Another concern with AJR's data was that soldiers' occupation put them at higher risk of mortality than actual settlers, which would upwardly bias the estimated effect of mortality as an exogenous variable. When involved in actual military operations, mortality rates from disease spiked. Albouy noticed in the data that soldiers were especially likely to be actively campaigning in colonies that had established so-called bad institutions, which could have easily led to an artificially high correlation between mortality and institutions in these areas. On the other hand, clergymen were more sedentary and determined where to establish bases the same way settlers did, by looking at factors such as access to clean water and climate severity (Nunn, 2012). Thus, missionaries lived more similarly to settlers and were better equipped to take precautions against disease than were soldiers, especially those on campaign. These characteristics should make my data a more representative substitute for lacking settler records.

In the end, after running several different models on my datasets, I conclude that these newly constructed mortality rates are not robust enough to be considered valid instruments for institutions. My process in arriving at the result will be explained later on in this paper.

LITERATURE REVIEW

Developmental economists have long struggled with the question of why some countries are poorer than others. It is obvious that the answer involves a wide variety of components. Preexisting literature has found statistically significant correlations between early institutions and current institutions (La Porta, Lopez-de-Silanes, Shleifer and Visny, 1999), and then current institutions and current GDP per capita. The, per se, right institutions provide incentives for economic growth through actions such as enforcing of property rights and opening a country's borders to trade (North and Thomas, 1976). The causal relationship between institutions and economic growth has been called into question since. Although institutions may seem to influence economic performance, the latter can also play a hand in what kind of institutions are constructed. The reverse causality issue here forces economists to find new ways of more accurately estimating this relationship. This concern for institutions being an endogenous regressor is the main motivation behind AJR's use of settler mortality as an instrument. This measure has also been positively linked to human capital growth, in replacement of institutions (Glaeser, Porta, Lopez-de-Silanes and Shleifer, 2004). Besides settler mortality, other economists have also found significant relationships between institutions and geographic location and crop fertility, which could also be used as prospective instruments (Easterly and Levine, 2002).

DATA

For comparison, AJR's data comes from Philip Curtin's *Death by Migration* and Vatican bishops records from Hector Guiterrez (1989; 1986). Curtin constructed mortality rates from medical records containing hundreds of thousands of individual observations from, mainly, French and British troops monitored between 1817 and 1836. Despite the large overall size of the sample base, these observations only appear in about 31 former colonies. About 20 of these are specified former African colonies, 4 other areas being general collections of African countries such as "French West Africa" (Curtin, 10). Curtin measured mortality as deaths per thousand. The record contains no information on birth or death years for soldiers, so Curtin could not construct age of death. Although the record does not specify age, soldiers tended to range from 17 to 50 years old. He estimated that troops stationed in the Western African colonies such as Sierra Leone and Senegal had that highest mortality rates, followed by troops within the West Indies, and Southern Asia. Troops within the Pacific Islands, Europe and the U.S. had the lowest rates of death.

Guiterrez compiled records of bishops stationed between 1604 and 1876, primarily in former Latin American colonies. Observations in this record are extremely minimal, often under 10 observations per region. Due to this obvious lack of observations, AJR prorated mortality numbers in order to match Curtin's measure of death per thousands. This data is meant to make up for lacking mortality rates for South America as whole in Curtin's military records.

In contrast, my panel data on Catholic bishops comes from records compiled starting from the 1st century until present (catholic-hierarchy.org). Figure 1 notes basic descriptive statistics of this data.

Variable	Obs	Mean	Std. Dev.
yearborn	11,436	1770.729	182.8986
ageordaine~t	7,732	25.94082	5.051255
ageappointed	11,352	54.47156	11.20403
agedied	10,895	70.96111	11.81803
yeardied	8,284	2088.984	1221.364

Figure 1

The set contains 17,022 bishops, of which 11, 436 have birth years and 10,895 have death years. In general, bishops averaged about 54 years of age at time of arrival in a colony. Just as important, 9,940 have reported locations of station.

Location	Freq.	Percent	Cum.
Central Africa	228	2.29	2.29
Europe	3,338	33.58	35.88
Mediterranean	83	0.84	36.71
North America	4,509	45.36	82.07
Pacific Islands	214	2.15	84.23
South America	759	7.64	91.86
Southern Africa	127	1.28	93.14
Southern Asia	506	5.09	98.23
West Africa	70	0.70	98.93
West Indies	106	1.07	100.00

Figure 2

Referring to Figure 2, most were assigned to places in Europe and North America, but with the use of general regions, there are still decent sample sizes for African areas such as Southern and Central Africa. In total, the dataset contains observations from 79 former colonies, 46 in Africa. AJR's dataset contains 64 former colonies, 40 of which hail from Africa. Yet, Curtin did not have records for a large majority of these African areas so this number is misleading.

The panel data on United Kingdom missionaries comes from historical documentation of the Society of the Propagation of the Gospel in Foreign Parts (SPG) (Pascoe, 1901). The SPG was a Church of England-sponsored organization that sent missionaries abroad to British colonies during the 1700s and 1800s. The main targets of their religious influence were European settlers, which makes the missionaries an especially representative group in the calculation of mortality rates.

Variable	Obs	Mean	Std. Dev.
startofsta~n	4,955	1861.619	40.43316
endofstation	4,955	1865.801	39.15181
birthyear	765	1833.467	38.68372
yeardied	613	1845.346	92.16764
yearsworking	4,955	7.106559	9.175098

Figure 3

As seen in Figure 3, of these 5,814 missionaries, only 765 have birth years and only 613 have death years. The average age of these missionaries upon entry into a colony was 29 years old. Despite the incompleteness of this aspect of the data, almost all have recorded locations of service. Looking at Figure 4, most traveled to North America and Southern Asia, however quite a few went to African regions as well. Out of 33 former colonies in this dataset, 19 are located in Africa.

ORIGIN AREA	Freq.	Percent	Cum.
Central Africa	2	0.04	0.04
Europe	197	3.53	3.57
Mediterranean	5	0.09	3.66
North America	2,223	39.83	43.49
Pacific Islands	594	10.64	54.13
South America	143	2.56	56.69
Southern Africa	874	15.66	72.35
Southern Asia	1,139	20.41	92.76
West Africa	26	0.47	93.23
West Indies	378	6.77	100.00
Total	5,581	100.00	

Figure 4

To tie the mortality estimates to current economic growth, this study uses the same datasets gathered by AJR. Unfortunately, many of these original, raw datasets are no longer publicly accessible, so numbers for institution measures and such are copied directly from AJR's per-country data file. Where the bishop or missionary file contains countries not found in AJR's, measures will be filled in based on available outside information.

For example, mortality is used as a regressor for European settlement populations, which is measured as ratio of Europeans in a colony's population in 1900. When this study's datasets had observations for the same countries as AJR's, this measure of European presence in a colony was copied directly. One exception was Equatorial Guinea, where several bishops were stationed. Luckily, in AJR's working paper, they mention the European settlement estimate for Equatorial Guinea despite not using it in their regressions, so this number is used in the bishop dataset.

Early institutions are measured by constraints on the executive in 1900, a seven-category scale with a score of 1 indicating no constraints against authority and a score of 7 indicating a type of egalitarianism among executors (Inter-University Consortium for Political and Social Research). There were no relatable data to substitute for this measure when the datasets did not align with AJR's, so countries such as Namibia and Rwanda are disappointingly dropped from regressions utilizing this institutional measure. The only exception to this is Botswana, which uses a statistic for constraints on the executive in 1990 found in AJR's "An African Success Story: Botswana" (2001).

Current institutions are measured by average protection against expropriation risk between 1985 and 1999, a ten-category scale with a score of 1 indicating highest risk of expropriation, considered "bad" institution. Estimates from The Global Economy database and the Polity IV Project serve as substitutes when average protection against expropriation risk is lacking for a particular country. When the scales for these alternative current institution measures differ from the original, the former are benchmarked to the later to ensure estimates do not become inflated.

Economic growth is measured by log GDP per capita in 1995 based on purchasing power parity. This data is still available through the World Bank. Although newer data is obviously available, this study will stick to the same measures used in AJR so as to maximize their results' comparability.

Average life expectancy in 1995 will be one of the controls for omitted variable bias when mortality is instrumented on current institutions and will increase robustness to the results. This measure will also come from World Bank data that is still accessible.

METHODOLOGY

This study will use a series of OLS regressions, based off of those run by AJR, to ultimately estimate the effect of colonial mortality as an instrumental variable for the effect of institutions on GDP per capita. The expected results here are still the same in that mortality rates will be positively correlated with institutions and that, in the end, low life expectancy rates will be a plausible predictor of a country having slow or little economic growth.

One of the most radical adjustments this study will make in replication of AJR's study is in the actual construction of mortality rates. As mentioned earlier, mortality is measured by number of deaths from disease per thousand per particular area. One of the biggest issues with this approach, as vocalized by Albouy, is the results' implications when data is lacking in numbers. Guiterrez' data did not contain at least a thousand bishops per area to observe deaths for. AJR attempted to bypass this problem with benchmarking, prorating the ratio of deaths to one thousand. This solution puts exceptional weight on what few observations were available to begin with, which allows for any extreme, abnormal cases of death to severely bias the mortality rate in a given location. Yet, there are little to no other more accurate ways of fitting minimal amounts of data to this death per thousands approach. Although the missionary and bishop datasets contain a significant number more observations per location than Guiterrez, this data, too, would have to be manipulated it in order to achieve this same mortality rate form.

Mortality calculations will be based on variation in life expectancies (please note that life expectancy and mortality rates are used interchangeably for the rest of the paper). This approach is more conducive to my panel datasets. However, the context is slightly different in the two types of data. Bishops rarely resigned, usually spending the remainder of their lives from the time of appointment in the foreign area. It is easy, then, to observe variation in remaining lifespan among many bishops assigned to different areas, when controlling for initial age differences at the time of relocation. While bishops could only return home by resigning, missionaries could essentially serve their time and retire back in their homeland. This explains why many missionaries in the dataset are recorded as having died at home. Yet, the use of life expectancy is still plausible in this dataset. Although some missionaries did not succumb to disease before their term expired, their experiences in a foreign environment have to have had some effect on their health following their return. Thus, these observations can still offer some insight into the likely mortality rates of average European settlers who might have stayed longer.

In order to construct average mortality rates for each location, there will be two separate regressions, one for each data set. For the bishops, age of death will be regressed on dummy variables for each recorded location served and some vector of fixed effects. Fixed effects will include age appointed (the age at which they relocated), and year of birth, which will control for the upwards trend in life expectancy. Fixing age will control for initial difference in age at the time of appointment.

The regression to calculate mortality will be slightly different for the missionary dataset because these records did not follow every person through the entirety of their life.

The regression to calculate average life expectancies for the bishops will be:

$$Age\ Died_i = \beta_i Station + \gamma_i Year\ Born_i + \delta_i Age\ Appointed_i + \varepsilon_i \qquad (1a)$$

in which β_i represents the variation in average life expectancy in each region and $\gamma_i$and δ_i are coefficients that control for variation in time of relocation and the positive trend in life expectancy. β_i will be

The same regression for the missionary data will look like:

$$Year\ Died_i =$$

$$\alpha_{1b} + \beta_i Station_i + \gamma_i Year\ Born_i + \phi_i Year\ Stationed_i + \pi_i Number\ of\ Stations_i + \varepsilon_I \quad (1b)$$

α_{1b} serves as the average mortality rate for North America. Just as the age appointed variable does in (1a), year stationed should also control for variation in time of relocation.

With regard to the dummy variables for location of station, colonies will be grouped together by general region as done initially by Curtin. Although this prevents analysis of individual effects, it does significantly increase the statistical power of each dummy. It is a better alternative to keeping dummies for individual colonies with an unjustifiably large amount weight on each observation. Grouping locations together will likely mute the effects of extreme mortality rates, however it is preferable to underestimate rather than overestimate the effects of disease. These mortality rates will then be reassigned to each country based their designated region, the same procedure used by AJR. It should be emphasized, however, that these are all countries actually observed in the missionary and bishop datasets, unlike the countries AJR added to their study based on their own justifications.

Once the estimated mortality rates have been distributed to every country in the datasets, the next series of regressions run will eventually connect them to current GDP per capita. They are the same used by AJR as there are no serious critiques of them. They are as follows:

$$S_i = \alpha_S + \beta_S \log(life\ expectancy)_i + \varepsilon_i \qquad (2)$$

$$C_i = \alpha_C + \beta_C S_i + \varepsilon_i \qquad (3)$$

$$R_i = \alpha_R + \beta_R C_I + \varepsilon_I \qquad (4)$$

(2) connects mortality rates to the size of European settlements in 1900 relative to total population in a given area, denoted as S_i. As almost all former European colonies were established by the late 1800s, settlers had at least decades afterwards to test the suitability of these environments before the time of the 1900 census collection. Thus, the relative size of Europeans who remained in these colonies by the turn of the century should reflect those that found these areas to be at least equally or even less life threatening than back in their mother countries.

(3) tests the theory that better colonial institutions were established where European settlements were higher. Early institutions, denoted as C_i, are measured as constraints against the executive.

(4) will connect early institutions to current institutions, denoted as R_i. Current institutions will be measured by ranking on the Risk of Expropriation Index. The last four regressions are meant to show the potential offered by mortality rates as an instrument—that they are correlated with settlement rates, which are correlated with early and current institutions. Once these calculated effects are shown to be statistically significant, the legitimacy of the instrument can be tested. As the main issue motivating the birth of AJR's

paper is the reverse causality between institutions and GDP per capita, the first stage and second stage regressions will be:

$$R_i = \alpha + \beta_i \log(life\ expectancy)_i + \vartheta_i + \varepsilon_i \qquad (5)$$

$$GDP\ per\ capita_i = \alpha + \beta_{IV} R_i + \vartheta_i + \varepsilon_i \qquad (6)$$

Note that ϑ_i refers to the use of average life expectancy in 1995 and dummy variables for the origin country of the colonizer as controls for mortality.

There are several mathematical and logical approaches by which this study can test the validity of mortality as an instrument. The instrumental relevance condition can be proven by the correlation between current institutions and mortality, which can also be shown by the coefficients on the variables in (2)-(4) being statistically significant. Logically, the exclusion restriction should also be satisfied as people today do not face the same mortality rates from the diseases that may have terrorized Europeans prior to 1900. Modern medicine has almost entirely eradicated the threats of illnesses such as malaria and yellow fever. Thus, these early mortality rates should not affect current economic growth. Empirically, the exclusion restriction will be tested by treating mortality as an exogenous regressor on GDP per capita and observing the estimated effect and significance level of the variable in this case. This regression will be:

$$GDP\ per\ capita_i = \alpha + \beta_i \log(life\ expectancy)_i + \vartheta_i + \varepsilon_i \qquad (7)$$

Running an overidentification test with mortality as an instrument for R_i should also test its validity. So long as diseases are uncorrelated with the error term—which is, again, logical—I can deem mortality rates, or life expectancy as legitimate just as AJR did.

For the sake of being thorough, I will also create mortality estimates using a Cox Proportional Hazard Model on top of the normal OLS regressions. The Cox Model is a type of survival analysis that will create death year estimates where they are censored in the data. Because almost all 5,000 of the missionaries have complete data for years stationed and duration, adding Cox estimated death dates will allow for the full use of the dataset in the estimation of life expectancies. The goal of this model is to calculate death dates based on actual observed survival rates of other missionaries within the same geographical region. It will treat the missionaries like patients in a medical trial, using year stationed as the time at which a person is put at risk of contracting a disease and death year as the time at which he succumbs to illness. Though the number of observations will increase, the logic behind this study still predicts that these new mortality estimates will be relatively the same to those constructed through OLS.

PRELIMINARY ANALYSIS

Constructing Mortality Estimates

As described in detail in the Methodology section, mortality will first be calculated based on the general region a missionary or bishop was stationed in. Figure 5 shows the output from (1a):

VARIABLES	(1) 1
Age Appointed	0.559***
	(0.0141)
Year Born	0.00213***
	(0.000136)
Central Africa	37.08***
	(1.421)
Europe	35.99***
	(0.820)
Mediterranean	38.18***
	(2.087)
North America	36.91***
	(0.835)
Pacific Islands	36.19***
	(1.511)
South America	34.47***
	(1.053)
Southern Africa	35.51***
	(1.808)
Southern Asia	34.86***
	(1.027)
West Africa	31.24***
	(2.308)
West Indies	33.96***
	(1.981)
Observations	5,726
R-squared	0.974

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Figure 5

Although the bishop dataset includes 17,022 observations, bishops who did not have records of both age of appointment and year of birth, or any record of location stationed were dropped from the sample. The R^2 showing that more than 97 percent of the variation in life expectancy is explained by (1a) is quite large considering how many omitted variables, such as bishop nationality, may also affect the dependent variable. Yet again, macro models commonly have R^2 of this magnitude.

Overall, all of the coefficients make sense. For example, year born being positive reflects the positive upward trend in life expectancy. Areas like the West Indies were home to yellow fever and malaria, the latter also being common in West African countries, cholera ran rampant in India, and Britain and France had drawn out battles with tuberculosis. These diseases explain the small size of the coefficients on the dummies in relation to the Mediterranean, Central Africa and North America. Though it may appear odd for the coefficient on Europe to be comparable to South Africa's, this takes into account the fact that each area is a combination of locations, which likely lowers the calculated effect of former African colonies in particular.

In terms of (1a)'s interpretive power, the calculated average age of death was around 71 years old for bishops assigned to North American stations. This is using the mean age of appointment of 54 years old and mean year of birth at 1770. However, considering how miniscule the coefficient on year born is, the difference in age of death when only looking at variation in time is almost not noteworthy. When comparing life expectancies, bishops assigned to areas like Algeria at the same age and time as those appointed to areas like Sierra Leone were expected to die almost seven years later than their counterparts. Considering all variables are significant at the 1% level, this model shows that mortality is highly correlated with geographic region.

The OLS missionary mortality rates had similar results. The output for (1b) is shown in Figure 6a:

VARIABLES	(1) OLS
Start of Station	0.964***
	(0.0152)
Number of Stations	9.130***
	(1.404)
Europe	63.46**
	(28.75)
Pacific Islands	70.94**
	(28.83)
North America	74.91***
	(27.36)
South America	65.89**
	(28.37)
Southern Africa	66.94**
	(28.74)
Southern Asia	66.98**
	(28.67)
West Africa	70.46**
	(29.65)
West Indies	63.95**
	(28.03)
Central Africa and Mediterranean	-
Observations	538
R-squared	1.000

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

Figure 6a

VARIABLES	(1) OLS Relative
Start of Station	0.964***
	(0.0152)
Number of Stations	9.130***
	(1.404)
Europe	-11.44***
	(2.523)
Pacific Islands	-3.963
	(4.293)
South America	-9.013***
	(2.217)
Southern Africa	-7.969***
	(2.252)
Southern Asia	-7.922***
	(1.955)
West Africa	-4.447
	(6.754)
West Indies	-10.96***
	(2.544)
Central Africa and Mediterranean	-
Constant	74.91***
	(27.36)
Observations	538
R-squared	0.934

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

Figure 6b

As a whole, many observations were dropped just as they were in (1a). The particular problem in this dataset, however, was lack of missionaries with both birth and death dates. Removing birth year increased the number of total observations from 93 to 538. Year stationed serves as the detrending variable and controls for differences in time of relocation. To deal lack of observations from particular regions, Central Africa and the Mediterranean are grouped together. Although the coefficient for this new variable is meaningless and will not be associated with countries within the dataset that belong to these regions, including the variable prevents the estimates of the other dummies from being skewed by some omitted variable bias.

The R^2 at 0.99 is of a similar size to that of (1b). One difference is the coefficients themselves. They are all still highly significant, albeit not as much as those in (1b), but the estimated relative effects are in a slightly different order. West Africa dropped in rank from the deadliest region in (1a) to one of the healthiest here. Europe now takes the number one spot, though it still has one of the higher death rates in the bishop data. Despite this, the estimates for South America, Southern Asia and Southern Africa maintain the same general ranking as in (1b).

The interpretation of the coefficients supports the underlying hypothesis. The positive sign on year stationed, once again, represents the positive trend in life expectancy. The number of different stations a missionary traveled to has a relatively large coefficient, signifying that every additional mission assigned increased one's life expectancy by a little over nine years. It is important to note that all of a particular missionary's posts were confined to the same general area, which eliminates concerns that the location dummies might double count when a person had more than one station. Using North America as an example, a missionary who began his work in 1861—the average for year stationed—in South Carolina and who, at some point, was transferred to Vancouver would be expected to live until 1886, a full 25 years after starting his mission. Despite the lack of age estimates here, estimating life expectancy based on time of entry into a colony is still clearly meaningful and reveals the same implications about the severity of these disease environments.

For sake of proving the validity of the variation in life expectancy among these regions, I also ran (1b) with North America as the omitted group. This result is shown in Figure 6b. The significance of these variables test the hypothesis that there is a real difference in mortality between regions. West Africa and the Pacific Islands were not significant here, implying that these regions' life expectancy are akin to North America's. Though the Pacific Islands does tend to have high life expectancy in (1a), this is a surprising result for West Africa which has a history of deadly diseases. Albeit, going back through the data reveals that of the few observations I had for West African countries, a few outlying missionaries survived within the country for over 20 years before dying. This could explain the high predicted life expectancy here. Despite this shortcoming in the dataset, it is better to underestimate than overestimate the effect of disease for this paper.

The Cox Proportional Hazard Model calculates life expectancy in the same way. The output from running this model is as follows in Figure 7:

VARIABLES	(1) Cox Proportional Hazard Model
Start of Station	0.994***
	(0.00108)
Number of Stations	0.460***
	(0.0435)
Europe	3.138***
	(0.975)
Pacific Islands	0.919
	(0.243)
South America	5.644***
	(1.173)
Southern Africa	1.499**
	(0.258)
Southern Asia	2.318***
	(0.299)
West Africa	1.890
	(0.789)
West Indies	1.473*
	(0.291)
Central Africa and Mediterranean	0
	(0)
Observations	4,309

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Figure 7

The main advantage of this model is the drastic increase in sample size from 538 in (1b) to 4,309. This model is interpreted through hazard ratios, a higher ratio signifying a higher risk of death. North America is treated as the omitted group. Once again, Central Africa and the Mediterranean are grouped together to diminish concern over omitted variable bias.

As seen by the p-values, most of these estimates are highly significant, the exceptions being Pacific Islands and West Africa. This result is unsurprising considering both of these regions were also insignificant in relation to North America in Figure 6b.

However, in addition to the insignificance of these variables, the relative ranking of mortality among these regions does not closely align with those of the OLS regressions for both datasets. Though the Pacific Islands does tend to have one of the highest life expectancies, these hazard ratios suggest that the West Indies and Southern Asia have similarly high rates. It implies that a missionary is three times as likely to die in Europe than in North America, but their likelihood of death is only slightly higher if they travel to Southern Africa instead of the latter.

In any case, I will run the remaining regressions with each of these mortality estimates to compare their effectiveness in predicting economic growth.

Using Mortality as an Instrument

As reminder, once I have calculated regional mortality estimates, I assign these numbers to all countries within the datasets that fall in the particular region. Running (2), (3) and (4) on both datasets produced the same relative results as AJR with highly significant coefficients, reinforcing the hypothesized indirect connection between settler mortality and current economic growth. Due to differences in the calculation of mortality rates, I cannot numerically compare them to AJR's, but they still imply that European presence in colonies positively influenced the construction of "good" institutions which persisted over and time and which promoted economic growth. These results are displayed in Figure 8 for the bishop dataset:

VARIABLES	(1) Constraints on Executive
European Population 1900	0.0480***
	(0.00744)
Constant	1.471***
	(0.223)
Observations	69
R-squared	0.383

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

VARIABLES	(1) Average Risk of Expropriation
Constraints on Executive	0.394***
	(0.114)
Constant	4.645***
	(0.352)
Observations	75
R-squared	0.139

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

VARIABLES	(1) European Population 1900
Bishop Mortality	532.0***
	(119.6)
Constant	-801.5***
	(183.9)
Observations	71
R-squared	0.223

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

VARIABLES	(1) log GDP per capita 1995
Average Expropriation Risk	0.331***
	(0.0434)
Constant	5.949***
	(0.264)
Observations	74
R-squared	0.447

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Figure 8

From here, the missionary and bishop regression output disagreed on the validity of mortality as an instrument. Figures 9 and 10 show the output from (5) and (6), respectively:

VARIABLES	(1) 1	(2) 2	(3) 3
Cox Mortality	-0.284		
	(1.585)		
Life Expectancy 1995	0.113**	0.119***	0.137***
	(0.0410)	(0.0336)	(0.0203)
British Colony	1.174	-0.537	1.142**
	(1.537)	(0.950)	(0.523)
French Colony	1.365	-0.657	0.978*
	(2.034)	(1.650)	(0.584)
Portuguese Colony	-0.795	-2.017	0.958
	(1.790)	(1.315)	(0.673)
Missionary Mortality		17.80**	
		(7.300)	
Bishop Mortality			-9.392
			(10.11)
Constant	-1.735	-75.51**	11.04
	(4.172)	(30.53)	(15.03)
Observations	29	31	77
R-squared	0.400	0.527	0.459

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Figure 9

VARIABLES	(1) OLS Missionary	(2) Cox Missionary	(3) OLS Bishop
Average Risk of Expropriation	0.228*	2.768	-0.180
	(0.129)	(13.40)	(0.371)
Life Expectancy 1995	0.0627***	-0.232	0.102**
	(0.0186)	(1.557)	(0.0474)
British Colony	0.0444	-3.378	0.455
	(0.275)	(18.46)	(0.512)
French Colony	-0.249	-4.046	0.277
	(0.484)	(20.68)	(0.466)
Portuguese Colony	-0.283	1.387	0.395
	(0.461)	(9.386)	(0.439)
Constant	2.804***	8.404	2.425**
	(0.825)	(30.51)	(1.151)
Observations	30	28	74
R-squared	0.811		0.545

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Figure 10

For the OLS missionary model, the output for (5) and (6) prove that the instrument is valid. This estimated mortality being significant at the 5% level in (5) shows that the relevancy condition is met. The positive sign on the coefficient reiterates that the higher the life expectancy, the less at risk the country was of having an extractive, or “bad” institution. Interestingly, current—1995 being considered current when AJR published their original paper—average life expectancy was the only other significant variable in the regression. It is meant to control for the current impact of disease in (6), however it may be picking up on the endogeneity issue of institutions here. Life expectancy today is usually positively correlated with GDP per capita as richer countries are better able to invest in medical technology advancements. The positive sign of its coefficient may likely be due to this correlation as disease should have no logical effect on the type of institution a country establishes—the only exception being the theory behind this paper. The controls for origin of the colonizer being insignificant reinforces AJR's findings that this aspect of colonization had no unique effect on the types of institutions as La Porta et all (1999) argued. Yet, including these dummy variables in the regression provides robustness to the assumption of this paper that the influence of mortality in a colony was high.

The results of (7) as seen below in Figure 11 prove that the OLS missionary instrument passes the exclusion restriction. Mortality is not reasonably significant with a p-value of 0.18, supporting AJR's hypothesis that diseases that affected people prior to 1900 have no current impact due to medical advances that have in many cases eradicated these problems.

VARIABLES	(1) 1	(2) 2	(3) 3
Missionary Mortality	-0.637		
	(0.463)		
Life Expectancy 1995	0.0850***	0.0807***	0.0774***
	(0.0128)	(0.0137)	(0.00733)
British Colony	-0.0776	-0.0696	0.286
	(0.345)	(0.511)	(0.200)
French Colony	-0.352	-0.272	0.128
	(0.593)	(0.677)	(0.222)
Portuguese Colony	-0.778	-0.798	0.246
	(0.479)	(0.595)	(0.249)
Cox Missionary Mortality		-0.718	
		(0.516)	
Bishop Mortality			1.834
			(3.724)
Constant	3.534***	3.616**	0.169
	(1.225)	(1.382)	(5.560)
Observations	32	30	75
R-squared	0.767	0.758	0.714

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Figure 11

These tests culminate in the output for (6), which instruments mortality on current institutions. The new estimate for current institutions is significant at the 10% level for the OLS missionary rates, again bolstering the validity of using mortality as an instrument. The estimated effect of current institutions increases from 0.11 in (4) to 0.23, more than a two-fold change. This increase in the coefficient of the endogenous regressor is a typical occurrence in instrumental variable regressions. Overall, the results of (6) seem to indicate that mortality is a viable instrument. This will later be retested with robustness checks.

The missionary mortality rates derived from the Cox Model did not produce the same conclusions. Cox estimated mortality was very insignificant with a p-value of 0.879 in

(5), indicating that it would be a weak instrument. This information alone is enough to justify the illegitimacy of using mortality, however (6) was also run to verify this information, which it did by estimating a p-value of 0.886 for the instrumented version of average risk of expropriation. Thus, Cox mortality estimates will not work as an instrument for current institutions and there is no further case for utilizing this model.

Strangely, the bishop data produced the same conclusions as the Cox Model estimates. Mortality was not statistically significant with a p-value of 0.365 when testing the relevancy condition in (5). The negative sign on the coefficient also seems to imply a negative correlation between life expectancy and institutions. This alone disproves the validity of the potential instrument as (5) implies that bishop mortality has a weak relationship with average expropriation risk. Running (6) regardless repeats this conclusion as the instrumental variable estimate is noticeably insignificant.

In Figure 10, the coefficient, -0.180, having a negative sign implies that those with less expropriation, the so-called "good" institutions like the U.S. and Canada, should have lower GDP per capita which is inaccurate.

The particular rejection of bishop mortality as an instrument while accepting the missionary mortality rates, at least for OLS, is perplexing. The bishop dataset contains more overall observations as well as over twice as many actual countries. The dataset is also more complete as life expectancy is calculated from bishop age, whereas incomplete records prevent this in the missionary dataset. Despite these differences, estimated mortality rates for each region had the same general rankings for both datasets. West Indies and South America had some of lowest estimated life expectancies, while North

America and Pacific Islands were near the top. Considering all of this and that the bishop mortality estimates were all significant at the 1% level in (1a), it should have been more likely that the bishop mortality be a valid instrument for current institutions if AJR's hypothesis does hold true. The failure of two out of the three models tested here places more doubt on the true viability of the OLS missionary mortality rates.

ROBUSTNESS

Several last checks will be performed on the OLS estimated missionary mortality rates in order to challenge the strength of this model. The controls used in (5) and (6) are the main tests of the robustness. The coefficient on mortality and the instrument estimate of average risk of expropriation did not change very noticeably whether average life expectancy in 1995 or origin of the colonizer variables were included or not. When included, the mortality coefficient in (5) changed from 17.8 (s.e. = 7.3) to 17.05 (s.e. = 7.31). Mortality remained statistically significant. For (6), including the control changed the instrumented coefficient from 0.23 (s.e. = 0.13) to 0.20 (s.e. = 0.11). Relatively, this last test of controls is definitely more noticeable, yet the instrument is still significant at the 10% level so adding this robustness does not necessarily change the conclusions here. Albeit, this robustness is called into question when adding a control for latitude, measured as the absolute value of distance from the equator scaled down to values between 0 and 1. Including latitude in (5) changes the coefficient on mortality to -15.9 (s.e. = 6.8), no longer making it reasonably significant. Adding latitude to (6) as an exogenous variable caused the p-value on the instrument to drastically increase from 0.077 to 0.655. Although latitude is never considered significant itself in these regressions, there is no reason for it to have such an affect of the significance of mortality. It is a commonly used control in AJR's regressions in reference to McArthur and Sachs' (2001), who used it as an instrument for the presence of disease. AJR found that, in every case, controlling for latitude had no real effect on the regressions' outputs or the implications. The fact that latitude did notably alter the interpretation of this study's regression output is a serious concern. It is the only test so far that has put the validity of these mortality rates into question.

Running overidentification tests are also important in questioning the validity of the instrument. Akin to AJR's strategy, the test consists of regressing log GDP per capita on the European settlement ratios and constraints on the executive, both variables serving as alternate instruments for average risk of expropriation. In addition, mortality is treated as an exogenous variable affecting log GDP per capita for comparison. This feature allows for testing if the regressions using the alternative instruments with and without mortality—serving, here as a normal exogenous variable—are significantly different. Based on the results of (7) discussed earlier, mortality should have no direct impact on current economic growth. Thus, regressions including mortality are not accurate. Showing that regressions with and without it are not differentiable proves that mortality is the only valid instrument. In the end, a simple Hausman test revealed p-values of 0.17 and 0.46, confirming that these regressions cannot be differentiated with confidence. Therefore, this test concludes that the OLS missionary mortality rates are a valid instrument for current institutions.

CONCLUSION

In recent years, developmental economists have begun to put more emphasis on the effects on types of institutions on the potential growths of economies. However, concern that type of institution is an endogenous regressor have forced economists to hypothesize alternate ways of calculating the true effect of institutions on economic growth. Among the hypotheses tested, AJR's use of settler mortality as an instrument for institutions has received the most recognition. AJR attempted to understand the origins behind the development of current institutions, especially in countries that were once colonized by Europeans. They postulated that where Europeans entering colonies faced high mortality rates they were more likely to leave behind institutions that did not promote local economic growth with these institutions mostly persisting over time. They showed that mortality rates were correlated with relative size of early European settlement populations, which were correlated with early institutions, which were correlated with current institutions. In the end, AJR concluded that settler mortality was a valid instrument. This study attempted to replicate this work using two new datasets that were more representative of settlers and addressed concerns raised by critics of AJR's model. Mortality rates from Catholic bishops and Protestant missionaries were calculated based on life expectancies, which were then used in the same regressions run by AJR for optimal comparison. Out of the three models run—OLS on the bishop data and both Cox Proportional Hazard and OLS on the missionary data—only the OLS missionary mortality estimates were significant when testing the validity of the instrument. Even still, the missionary estimates did not pass all of the robustness checks. As there to truly prove the validity of a potential instrument, this study is hesitant to agree with AJR's conclusion that mortality is valid. There is simply no explanation as to why only one particular model of the missionary data seemed to pass the validity tests but the bishop data did not.

Nevertheless, these findings are an important addition to the discussion of potential instruments for institutions. This study is the first time AJR's hypothesis has been tested with different data. Ideally, more prospective data composed of larger samples of Europeans similar in nature to settlers for every former colony observed will surface and this hypothesis can be further tested.

Work Cited

Acemoglu, D., Johnson, S., & Robinson, J.A. (2000). *The colonial origins of comparative development: An empirical investigation* (No. w7771). National bureau of economic research.

Acemoglu, D., Johnson, S., & Robinson, J. A. (2002). *An african success story: Botswana*.

Albouy, D. Y. (2008). *The colonial origins of comparative development: an investigation of the settler mortality data* (No. w14130). National Bureau of Economic Research.

Cheney, D. (n.d.). The Hierarchy of the Catholic Church. Retrieved December 11, 2015, from http://www.catholic-hierarchy.org/

Curtin, P. (1989). *Death by migration: Europe's encounter with the tropical world in the nineteenth century.* Cambridge [England: Cambridge University Press].

Easterly, W., & Levine, R. (2003). Tropics, germs, and crops: how endowments influence economic development. *Journal of monetary economics*, *50*(1), 3-39.

Glaeser, E. L., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2004). Do institutions cause growth?. *Journal of economic Growth*, *9*(3), 271-303.

Gurr, T (1997). *Polity II: Political Structures and Regime Change, 1800-1986.* Unpublished manuscript, University of Colorado, Boulder.

Gutierrez, Hector (1986). “La Mortalite des Eveques Latino-Americains aux XVIIe et XVII Siecles,” *Annales de Demographie Historique*, 29-39.

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer and Robert W. Vishny (1999). “The Quality of Government,” *Journal of Law, Economics and Organization,* 15, 222-279.

McArthur, John W. and Sachs, Jeffrey D. “Institutions and Geography: Comment on Acemoglu, Johnson and Robinson (2000).” National Bureau of Economics Research (Cambridge, MA) Working Paper No. 8114, February 2001.

North, D.C. and R.P. Thomas (1973). *The Rise of the Western World: A New Economic History,* Cambridge University Press, Cambridge UK.

Nunn, N., Akyeampong, E., Bates, R., & Robinson, J. A. (2011). Gender and missionary influence in colonial Africa. *African Development in Historical Perspective*.

Pascoe, C. (1901). *Two Hundred Years of the S.P.G.: An historical account of the Society…1701-1900. Based on a digest of the Society’s Records. By C.F. Pascoe.*

Find Data. (n.d.). Retrieved March 14, 2016, from http://www.icpsr.umich.edu/index.html