Hither Thou Shalt Come, But No Further: Reply to "The Colonial Origins of Comparative Development: An Empirical Investigation: Comment"
Online Appendixes and Appendix Tables

Daron Acemoglu — **MIT**

Simon Johnson — **MIT**

James Robinson — **Harvard**

March 2012.

Appendix A: Anderson-Rubin Confidence Intervals

To construct Anderson-Rubin (AR) confidence sets without clustering (with spherical errors), we first calculate the Anderson-Rubin test statistic and then invert it by solving a series of quadratic polynomial inequalities. The solution to these inequalities will correspond to a finite interval, the union of two infinite intervals, the whole real line, or an empty set. A fast, accurate algorithm for solving these inequalities has been developed and programmed for Stata by Anna Mikusheva and Brian Poi, in the form of the *condivreg* module. This approach is faster and more accurate than inverting the AR test statistic using a grid test, which performs a series of hypotheses tests $H_0 : \beta = \beta_0$ where β_0 belongs to a grid (Mikusheva 2010). More details about the algorithm and its STATAS implementation can be found in Mikusheva and Poi (2006).

Mikusheva and Poi's algorithm is applicable only to the spherical case. To calculate the clustered AR confidence sets (with non-spherical errors), we must invert the AR test statistic through grid testing. We do this by using the *rivtest* module in STATA which is documented in detail in Finlay and Magnusson (2009). In particular, as outlined by Chernozhukov and Hansen (2008), we first regress a transformed dependent variable, $Y - X\beta_0$ on the instrument, Z, where X is the endogenous regressor: $Y - X\beta_0 = Z\alpha + \epsilon$. We then test that $\alpha = 0$ (which is implied by $\beta_0 = 0$) using a conventional robust covariance matrix estimator. Finally, the AR test statistic is inverted using a grid test. For the results reported here, we searched an evenly spaced grid containing 1600 points (the maximum allowed by the Stata program) on an interval 25 times the width of the Wald confidence interval. We checked robustness to using wider and narrower intervals, as well as to searching on an evenly spaced grid ranging from -20 to 20, and in all cases the estimates changed very little from those reported here.

Appendix B: Information on Disease Ecology

The relevant literature is large, beginning with Lind's *Diseases in Hot Countries,* the first edition of which was published in 1768 and which helped form early perceptions. Hirsch (1888) provided the comprehensive 19th century compilation and assessment of knowledge; Volume 1 of his *Handbook of Geographical and Historical Pathology* covers "Acute Infective Diseases". As this was written before the transmission of disease by mosquito was properly understood - first published in German in 1881, its three volumes are the result of 25 years of work - it is particularly valuable as a window on contemporary understanding of Medical Geography. Our core assessments below are based primarily on the modern equivalent to and update of Hirsch, which is *The Cambridge World History of Human Disease*, edited by Kenneth F. Kiple (1993).[1] This volume contains more than 1,000 pages on almost all known human diseases, including sets of chapters on "The Geography of Human Disease" (by region), "The History of Human Disease in the World Outside Asia," and "The History of Human Disease in Asia." We checked the assessments there with H.O. Lancaster's *Expectations of Life: A Study in the Demography, Statistics, and History of World Mortality*, which contains chapters by region.

As a way to check that these general assessments fit with the detailed geographes in question (including with modern borders), we also checked a series of large world maps published in 1951 by the American Geographical Society, from *The Geographical Review*, Vol.41, 1951. This provide information on the incidence of infectious disease in every country of the world. Of particular value for our purpose is the "Distribution of Malaria Vectors" (American Geographical Society, 1951a; also useful are American Geographical Society 1951b, 1951c, and 1951d), which shows the type of mosquito "Species of Anopholes") present in every country - this has an important effect on the potential prevalence of the more serious forms of malaria - as well as the distribution of malaria parasites. This map provides references, by country, with most of its sources dating from the 1920s, 1930s and early 1940s. This graphic was obviously drawn after the early colonial period that is our focus here, but in most of Africa, Asia, and Latin America infectious disease was not conquered until the 1940s. Specifically, relatively little progress was made towards eradicating malaria, yellow fever or other mosquito borne disease before the 1940s, although there were advances in lowering infection rates for Europeans even in places that had previously been unhealthy for them (see Acemoglu and Johnson,

[1] Kiple's team has the benefit of hindsight and contains today's leading medical historians but might be considered somewhat distant from events and perceptions of the nineteenth century. Between Hirsch (1888) and Kiple (1993) there is Clenow (1903) - whose volume benefits from the medical advances at the end of the nineteenth century but who is still close to the major mortality events of that century. There is no indication in Clenow (1903) that our assessments based on Kiple (1993) are off the mark on anything that matters for our analysis.

2007). As Curtin (1989) discusses the importance of yellow fever epidemics in accounting for relatively high mortality during the 19th century, we also make use of the American Geographical Society's map showing the "Distribution of Dengue and Yellow Fever."[2] We check these maps against the latest available information in the medical geography literature, focused on epidemics, in the form of Hoff and Smith (2000) and Cliff, Haggett, and Smallman-Raynor (2004).

R.R. Kuczynski's three volume *Demographic Survey of the British Colonial Empire* also helps as a further cross-check. Published in 1948, this contains a great deal of the available historical demographic and public health information by country - particularly for African colonies. His three volumes cover West Africa (Volume I), South Africa, East Africa, Mauritius and Seychelles (Volume II), and West Indian and American Territories (Volume III). The lack of serious progress through the 1940s against infectious disease in many places were still colonies also makes twentieth century conditions in some places quite relevant for assessing the pattern of disease ecologies in the 19th century.

In terms of specific diseases, Curtin emphasizes the importance of differential incidence of malaria - and variation in the types of malaria across regions (also see Bruce-Chwatt and Bruce-Chwatt 1977). We use the updated version of Curtin's recommended text on this issue, which is Bruce-Chwatt's Essential Malariology, Third Edition (Gilles and Warrell 1993). We supplemented this information with Desowitz (1991) and Bradley (1992). The most useful sources in the historical record of malaria eradication efforts are League of Nations Health Organization (1932), Expert Committee on Malaria (1947) and Bulletin of the World Health Organization (1954). Conybeare (1948), Stolnitz (1955), Davis (1956), Caldwell (1986), and Preston (1980) are also helpful - on the point that malaria rates declined only from the 1940s so our sources from the early 1900s and later (up to and including Kuczynski 1948 and the American Geographical Society 1951a) are relevant for assessing 19th century disease ecologies.

We also looked carefully through the demographic and public health literature for specific articles on health conditions before 1940 that would shed light on local disease ecology - particularly anything that would suggest neighboring countries did not share the same disease ecology. This includes searching journals such as *Demography*, *Population Studies*, *Population and Development Review*.

This literature allows us to look carefully for any neighbor "anomalies", i.e., conditions under which neighboring countries would not share the same disease ecology. Perhaps the leading example is the disease barrier provided by the Sahara desert. In this regard, Albouy makes an important and elementary mistake when he argues that just because some West African countries border some North African countries, these two sets of countries might have a similar disease ecology - and therefore similar mortality rates for Europeans in the 19th

[2] For more on the role of yellow fever in high European mortality before 1900, see Oldstone (1998), Chapter 5, "Yellow Fever".

century. Patterson (1993) and Kuhnke (1993) - adjoining chapters in Kiple (1993) - make it very clear that this is not the case: West Africa and North Africa have fundamentally different disease ecologies, particularly with regard to the conditions for mosquitos, the vectors for key diseases that killed a high percentage of exposed Europeans before 1850.

The actuarial literature, which developed rapidly after about 1850, provides a useful cross-check - typically based on decades of experience for particular life insurance companies. Institute of Actuaries (1851-52) reports insurance rates used by "London Offices" in mid-century. Meikle (1876) assesses life insurance experience for Europeans outside of Europe. Hunter (1907) provides a review of life insurance experience in the last decades of the 19th century around the world - and suggests a classification of countries by mortality category.[3]

The remainder of this sub-section reviews each instance when Albouy drops our data.[4]

Australia

In AJR (2001) we assigned the New Zealand rate (8.55 per 1,000) to Australia, based on Marshall (1993) and the broader disease ecology literature. Curtin (1989) has a lengthy discussion of health conditions in New Zealand, including why exactly it has always been malaria free. In our assessment, Australia shared those characterisitics. Albouy finds this unconvincing and drops Australia from his core dataset.

However, in writing AJR (2005), we found that Tulloch (1847, p.253) reports mortality prior to 1836 in New South Wales and Van Diemans Land (Australia) as 14 per 1,000, with about the same rates in 1844-45. Albouy now accepts this rate for his extended mortality dataset.

But we also pointed out that on the first page of his introduction to part I of Army Medical Department (1840), Tulloch argues that "more than a fifth part arose from violent or accidental deaths" and "Thus the mortality from disease alone could have amounted to little more than one per cent annually, being lower than in any other Colony, except the Eastern Provinces of the Cape of Good Hope, to which the climate of Australia is in many respects similar." In other words, Tulloch puts deaths from disease in Australia at 10 per 1,000.

To err on the conservative side, we used the rate of 14 per 1,000 in our "Tulloch" revised dataset, with results shown in Tables 3A and 3B of AJR (2005), so that we could examine whether our initial assumption of 8.55 per 1,000 makes any difference - and it does not. But in any case there is no defensible rationale for dropping Australia - our data about 19th century mortality in this country are almost as good as our data for Europe.

[3] The life insurance data has a lower upper bound than the original data we used in AJR (2001). The supports the idea of a cap on maximum mortality rates, as discussed above. Acemoglu, Johnson, and Subramanian (2011) examine this issue further.

[4] We do not deal here with points that he raised before but that he has now withdrawn from his regression analysis. See AJR (2005, 2006, and 2008) for those details.

Singapore

AJR (2001) used the Straits Settlements estimate in Table 1.1 (Curtin, 1989) for both Malaysia and Singapore.[5] Albouy accepts our estimate of the mortality rate for Malaysia but discards the data from Singapore. But, as we reported in AJR (2005), there is strong qualitative confirmation of our coding.[6] A committee of the Statistical Society of London (1841) (a source for Curtin 1989, Table 1.1), wrote "On the whole, the town is distinguished by its salubrity; and it is a remarkable fact, that notwithstanding that the settlement is surrounded by marshes, and is exposed to many of the causes which are usually supposed to create malaria, malignant remittant fever has not been known there since its formation" (p.139)[7] There is no mention of malaria or yellow fever in the early medical history of Singapore; see Kiat (1978).

Furthermore, Albouy allows Singapore in his "revised mortality" series, but with the same rate it has in the AJR base sample. Dropping Singapore from his core dataset is therefore contradictory as well as making no sense.

Guyana

Tulloch (1838a) reports a mortality from disease rate in British Guiana of 84 per 1,000 over 1817-36 (pp. 131 and 133).[8] Previously we used the rate from French Guyana (32.18 per 1,000; Table 1.1 in Curtin 1989), so this direct estimate is presumably preferable and we use it in our robustness series. The public health literature on British Guyana itself does not indicate any anomaly that would suggest its disease ecology is different from that of French Guyana (Roberts, 1948; Mandle, 1970). Albouy drops Guyana from his base sample.

Hunter (1907, p.401) puts "British, Dutch, and French Guiana" in the same mortality category ("tropical").

Dominican Republic and Haiti

For the Dominican Republic and Haiti we used the Jamaica mortality rate of 130 per 1,000. Albouy drops these datapoints. The extensive disease ecology and historical literature for the Caribbean distinctly indicates that the Dominican Republic and Haiti were on the high side of typical mortality.

Institute of Actuaries (1851-52) reports an extra premium for life insurance in all the West Indies of 100 shillings. Hunter (1907, p.401) explicitly puts Haiti in the same category as other Caribbean countries, including Martinique and Guadeloupe (mortality rates of 112.18 and 106.87 respectively in Table 1.1 of Curtin 1989.)

[5] The estimate is from Penang; both Malacca and Singapore were part of the Straits Settlement.

[6] The authors of this report co-operated with Tulloch (Statistical Society of London, 1840, p.114)

[7] They also say, p.139, that in other parts of the island, "it is stated that fevers and dysentery are frequent." This supports AJR's contention that Singapore was healthier than its immediate surroundings.

[8] The average strength of the force was 884 (Balfour, 1845, p.201).

Bahamas

For the Bahamas, we used Curtin's estimate for the Windward and Leeward Command (85 per 1,000). Tulloch (1838b, p.229) reports that with an average strength of 27, the Bahamas had 102 verified deaths from disease over 20 years (1817-37), which is an average annual death rate of 189 per 1,000.[9] Most of the mortality occurred during epidemics in 1819 and 1823 and Tulloch attributes this high death rate to the unfortunate location of one fort, which was particularly vulnerable to yellow fever. We use the rate of 189 per 1,000 in our revision.

Albouy drops the Bahamas from his core sample.

Meikle (1876, p.277) assesses mortality in the Bahamas as very similar to that in Mauritius (which is 30.5 per 1,000 in Table 1.1 of Curtin 1989.)

Hong Kong

For Hong Kong we used the China Field Force rate for the British army in 1860, from Table A8.2 of Curtin (1998), which is 14.9 per 1,000. Albouy discards Hong Kong, regarding it as completely missing data for his core sample. For AJR (2005) we looked at the historical record and found legitimate discussion over which estimate to use for Hong Kong as various numbers are available.

The death rate for "White troops" in China in 1859 was put at 41.93 per 1,000 by Balfour (1861) - 59.35 per 1,000 including invalids who died on the way home (not usually included in early mortality estimates) - and 52.04 in Southern China in 1860, which includes invalids left in Hong Kong (Army Medical Department 1862). Jannetta (1993) and Leung (1993) give no indication that Hong Kong had a disease ecology that was significantly different from nearby parts of mainland China. And we can find no indication in American Geographical Society (1851a) or other maps that Hong Kong was any kind of disease anomaly.

However, there is a higher mortality estimate for European soldiers in Hong Kong from Tulloch's writings. On p.254, Tulloch (1847) reports an average ratio of mortality per 1000 of strength, 1842-1845, to be 285.[10] This is not a long average, as in the rest of Tulloch's work, but we still take this rate for our robustness series in AJR (2005). It is also not certain that all these deaths are from disease, but it does fit with Cantlie's negative assessment of Hong Kong (Cantlie 1974).

Tulloch's very high mortality estimate does not fit the fact that the British and French used Hong Kong as a gathering point for the China Field Force in 1860. Why would they have done this if the place were known to be so unhealthy? As Graham (1978, p.386) says, Kowloon was "an apparently healthy site for a barracks or camping ground...". Was there perhaps a big

[9] Total deaths were 107, of which five were "causes not known." If we use total deaths, the mortality rate would be 198 per 1,000.

[10] This is in an article that is not cited by Curtin, and which we overlooked in writing AJR (2001) - we reported the data in AJR (2005).

mortality difference between Hong Kong island (presumably covered by these statistics) and the Kowloon Peninsula (where the troops mustered)? Select Committee (1866) suggests part of the answer – troops were sent to Hong Kong when already sick, thus raising the measured mortality rate. However, the evidence and proceedings of this committee suggest there was some malaria in the area at that time.

In Institute of Actuaries (1851-52, p.169) all of China is in the same category of "extra premium" for life insurance (60 shillings for civilians and 80 shillings for military). Hunter (1907, p.401) puts the entire Chinese Empire, "south of 30 degrees North latitude", in the same mortality category ("tropical").

When there are varying estimates, we can average or use the alternatives or find some other way to combine the data. But to discard the observation completely, i.e., to treat it as " missing", is not appealing.

Pakistan

All our data for South Asia came directly from Curtin (1989, Table 1.1). We assigned the available rates to modern countries as follows: Bangladesh from Bengal (71.41), Madras for India (48.63), and Bombay for Pakistan (36.99 per 1,000). We also used the rates from Ceylon for Sri Lanka (69.8).[11] These data are all from before 1838. Albouy disputes - and drops - only the data for Pakistan.

But the British perception of health conditions in modern day Pakistan and nearby regions was very close to their view of mortality around Bombay. Bhardwaj (1993) gives no indication that the area covered by modern Pakistan is significantly different from other parts of the northerly-western region of what is now India.

For Pakistan we have gone carefully through the extensive British reports on military mortality in 19th century India. In our assessment, the British area of operations close to and including modern Pakistan, the expected mortality rate was at or close to what we included in our original series.

In Institute of Actuaries (1851-52, p.169) all of India is in the same category for "extra premium" on life insurance (60 shillings for civilians and 80 shillings for military). In Hunter (1907, p.401), all of (then-British) India is placed in one mortality category ("tropical")

Morocco

In AJR, the mortality rate for Algeria (78.2 per 1,000) is assigned to Morocco; while Tunisia was 63 per 1,000 and Egypt was 67.8. Institute of Actuaries (1851-52) puts the extra premium for Europeans traveling to Morocco at 40 shillings, which is the same as for an "Eastern Tour" that includes Egypt and other parts of the Middle East;.

[11] See Army Medical Department (1841, p. 8) for the original estimate and more detail. Mortality may have been lower 1820-26, but the data are not strictly comparable.

West Africa[12]

Our West African estimates were all from Curtin. From Curtin (1989, Table 1.1) we took data on early soldiers in Sierra Leone and Senegal, and from Curtin (1998) we used data on soldiers and small expeditions somewhat later in the 19th century (for Gambia,[13] Gold Coast/Ghana,[14] Mali/French Soudan,[15] Nigeria[16]). Specifically, we took data from expeditions with a few hundred soldiers on short West African expeditions (travelling on steamers or on mules); these were essentially peacetime experiences, with reported deaths almost all from disease. Curtin (e.g., 1990) emphasized an important downward bias from using data later in the 19th century, as militaries became better at managing mortality during short expeditions during the 19th century, so we stayed away (as much as possible) from estimates after 1850.[17] In the robustness checks of our NBER working paper, however, we did check our results using longer averages of African data; our main results were unchanged.

For Africa, we assigned mortality based on the literature on disease ecologies - erring on the side of using relatively low mortality rates and not those conspicuously from epidemics. This assignment is supported also by the life insurance literature. We have always emphasized that the data for some parts of Africa are less reliable than for other regions - and this has motivated our robustness checks without Africa (see AJR 2005 and the tables in this paper; this was also the focus of AJR 2006). But disregarding all African data completely is an inappropriate approach given the extensive available information. There is without doubt a great deal of measurement error in the African data but there is also much information about early European mortality in that region - and Europeans at the time were well aware of this.

We assigned mortality rates to countries that were part of the same colonial area or neighbors in the cases of Niger (from Haut-Senegal-Niger), Burkina Faso (from French Soudan), Guinea (from Sierra Leone), Cote d'Ivoire and Togo (from Gold Coast/Ghana).[18] Albouy

[12] There are 11 West African countries in AJR: Burkina, Cote d'Ivoire, Ghana, Guinea, Gambia, Mali, Niger, Nigeria, Senegal, Sierra Leone, and Togo.

[13] The Gambia data are "soldiers on the Gambia in 1825" (Curtin, 1998, p.10). "In this case, between May 1825 and December 1826, fevers killed 279 British soldiers out of a force that was seldom more than 120 and often as low as 40." For confirmation this was a peacetime experience, see the original source, Army Medical Department (1840, p.13).

[14] The Ghana (partly the Cape Coast Command) estimate of 668 was for troops 1823-26 and officers 1819-36. The original number is in Army Medical Department (1840, p.19). For confirmation, see Balfour (1849, p.38).

[15] The Mali expedition (specifically to Logo in 1878) included 434 Europeans and 225 Africans, travelling by steamer (with a march of 10 miles at the end); 49 percent of the Europeans died in less than two months (Curtin, 1998, pp.80-81). In campaigns in the French Soudan, under the direction of General Gallieni, soldiers rode on mules (Reynaud, 1898, p.150).

[16] The Nigeria expedition in 1841 had 159 Europeans on three steamers; "the longest time any of the steamers spent on the river that year was just over two months" (Curtin, 1998, p.21). The Ghana rates were from a longer intervention, 1824-26 (Curtin, 1998, p.18).

[17] "A mortality revolution had nevertheless taken place during the nineteenth century in tropical Africa as it had in Europe" (Curtin, 1990, p.69). From Army Medical Department (1840, e.g., p.22) it is clear that the early mortality estimates for West Africa are underestimates as they do not include deaths of soldiers once they had been "invalided" home.

[18] Albouy complains (p.9) that we assign a rate of 400 (Curtin, 1998, p.85) from Mali to Niger, but this is

drops these five countries.

But according to Curtin's *Image of Africa* (Curtin 1964, e.g., chapter 3), the Europeans had a clear and negative view of mortality throughout West Africa. There is no hint in the historical record that any of these places were regarded as potentially more healthy. From Curtin 1964, p.71, the discussion is all about West Africa.

Speaking of the late 18th century (i.e., before Tulloch's pioneering statistical work), Curtin (1964, p.71) articulates the perceptions of Europeans this way, " West African mortality figures were not widely publicized or given statistical precision, but the region's general reputation for having a "deadly climate" rested on a basis in fact. Somewhere between 25 and 75 percent of any group of Europeans newly arrived on the Coast died within the first year. Thereafter, the death rate was much less, perhaps on the order of 10 per cent per annum, but still substantial. Any European activity demand a price in European lives that was not only intrinsically high, but considerably higher than the cost of similar activity in the West Indies or South Asia. Slightly later calculations of military mortality over twenty years show a loss of 483 per thousand mean strength among European troops in West Africa, against only 78.5 per thousand in the West Indies. Civilian life insurance premiums charged by British firms for different tropical regions tell a similar story: European mortality was roughly four times as high in West Africa as it was in India or the West Indies." The early mortality rates reviewed in Bruce-Chwatt and Bruce-Chwatt 1977, pp.43-50) are entirely consistent with this assessment.

The explorer Richard Burton described Lagos Government House in 1863 as a "corrugated iron coffin or a plank-lined morgue containing a dead Governor once a year" (Bruce-Chwatt and Bruce-Chwatt 1977, p.47). Bruce-Chwatt and Bruce-Chwatt (1977, p.47) report that "the annual death rate for these high officials [Governors of Sierra Leone] was around 200 per 1,000". Kuczynski (1948, volume 1, pp.40-153) provides more details on the mortality rates of Europeans in Sierra Leone and other parts of British West Africa during the 19th century; this is completely consistent with the work of Tulloch and Curtin.

Albouy is also concerned about our assignment of the estimate of 280 from Curtin (1998), p.238, Table A8.1, for "French Soudan." The term French Soudan is ambiguous, as Albouy points out. As far as we know from Curtin and Reynaud, these were minor campaigns, with little fighting, mostly in present day Mali. Alternative assignments to Mali and its neighbors (e.g., assigning our original Mali estimate to neighbors, or using the estimate of 400 per 1,000 on p.85 of Curtin, 1989) make little difference to our results.[19]

assignment to a neighbor with the same disease ecology.

[19] Our original Mali estimate was very high, so we were reluctant to use this for all neighbors. But using this would be a reasonable robustness check. Assigning the rate of 2920 to Niger, Burkina Faso, and Cameroon (the last not a neighbor, but close and a neighbor of Nigeria, which has a similarly high rate in our base data), gives a parameter estimate of -0.54, with a clustered standard error of 0.13, without other covariates. If we assign 400 to Mali, Burkina Faso and Cameroon (Niger is already at 400 in our base data), the coefficient is -0.62 and the standard error is 0.18. In the first case, the coefficient falls slightly in absolute value, but the standard error also declines, and in second case there is almost no change (compare with column 1, Table 1A in AJR 2005).

In Institute of Actuaries (1851-52, p.169), the extra premium for life insurance in West Africa is 160 shillings in Senegambia and 120 shillings in the rest of the region. According to this source, these were the highest mortality places in the world for Europeans.

Chapter VII of Hirsch (1881, section 60, pp.198-202) identifies all of West Africa as an intensely malarial area. His sources on more inland West Africa were more limited - but Patterson (1993) is clear that this is the same disease ecology (unlike, for example, North Africa, which is quite different).

Central Africa

From Central Africa, Albouy drops Angola, Cameroon, Gabon, Tanzania, Uganda, Zaire. For Central Africa mortality estimates were particularly hard to come by and in AJR (2001) we proceed cautiously by relying on two relatively conservative numbers. First, we assigned a mortality rate of 280 from French Soudan to Angola, Cameroon, Chad, the Central African Republic, Gabon, and Uganda. This estimate is from Curtin (1998), p.238, Table A8.1. The disease ecology literature suggests there was less yellow fever in Central Africa than in West Africa (see American Geographical Society 1951e), but still significant amounts of falciporum malaria.

Second, from Curtin we had estimates of mortality rates for Africans working away from their homes, for Congo and Kenya. We took the highest observed values of these rates to represent a minimum for Europeans in those places. The Kenya rate (145) was assigned to Tanzania, and the Congo/Zaire rate (240) was assigned to Congo-Brazzaville and Zaire; these rates and the underlying source (Curtin et al. 1995) was stated clearly on p.33 in our NBER working paper.

Kiple (1984), Chapter 10, discusses the relative mortality rates of Europeans and Africans in places with malaria and other tropical diseases. The data, from the same underlying source as Curtin uses, suggests that before tropical medicine improved in the mid-19th century, the death rate for Europeans would be 2-3 times the death rate for Africans (and sometimes higher – see Tables 4 and 5 on pp.170-171, Kiple 1984).[20]

Sprague (1895, p.69) writes in the life insurance literature, "The Central Congo district has such a bad name that the mortality among Europeans resident there is said to be about 25 per-cent." According to Institute of Actuaries (1851-52), all of Central Africa is in the same high category of mortality for Europeans - requiring an extra premium on life insurance of 120 shillings. Hunter (1907, p.402) puts West Africa and Central Africa in the same mortality category ("higher than tropical scale"). American Geographical Society (1951e) shows all of West and Central Africa to be in the same endemic yellow fever zone. American Geographical

There is a similar pattern in other specifications.

[20] Curtin (1998, Table 1.1, p. 8) reports deaths from disease in the Sierra Leone Command, 1816-37, as 26.5 for Africans and 478 for Europeans.

Society (1951a) shows West and Central Africa share the same prevalence of *anopholes gambaie* - the primary vector for falciporum malaria.

Hirsch (1881, p.199) indicates that the Congo Coast was somewhat more healthy - with less malaria - than West Africa, but considerably less healthy than southern Africa. Patterson (1993) does not mention any significant difference in disease ecology between West and Central Africa.

Dropping Congo/Zaire makes no sense - as this is the area covered by our data and other available information. Albouy does not drop Kenya, for which we also had similar data directly.

Appendix C: "Campaigns"

We have reexamined the historical record for every one of our observations. While we do not claim to have established definitively whether there was or was not significant campaigning in each episode covered by our settler mortality estimates, here are some blatant examples of miscoding "campaigns" and "barracks" in Albouy.

We also indicate whether we recode the observation as campaign in either our minimal or extended recoding (note: all countries recoded as campaign in the minimal recoding are coded as campaign in the extended recoding). The choice of whether a country is in the minimal or extended recoding category is somewhat arbitrary, but doesn't make a significant difference to our results.

Jamaica - 1817-1836, this period includes the largest slave uprising in Jamaica's history known as the Baptist War in 1831. So there is fighting and campaigning during the period under consideration. This war is discussed in every book on Jamaican history; a much cited academic article is Reckord (1968). In our extended recoding, Jamaica is coded as a campaign.

Sri Lanka - Curtin has this number from 1817-1836. The Dutch had controlled the whole of the Island except for the Kingdom of Kandy; the British fought a series of wars after 1803 to annex this. The 3rd Kandyan War, took place 1817-1818, which is inside the period covered by Curtin. This war was big and it is discussed in every history of Sri Lanka. For instance, Peebles (2006, p. 50) notes that 1,000 British troops died. In our extended recoding, Sri Lanka is coded as campaign.

Malaysia and Singapore - these data are from the Straits Settlement 1829-1838. In 1831-32 the British fought the Naning War. Mills (1966) describes this in Chapter 7 pp. 115-128 and notes on page 115 that there was 9 months of campaigning. The war took place near Melaka, part of the Straits Settlement, for which we have data. Mills says that Indian soldiers were involved but he also continually talks about British forces. This is a war with British forces campaigning, right in the middle of the period Curtin defines. In our extended recoding, Malaysia and Singapore are coded as campaign.

Hong Kong – 1860 China field force. This number comes from Table A8.2 in Curtin

(1998, p. 239) and in this table this is described as a "campaign." Albouy must have misread this table, and in our minimal recoding Hong Kong is coded as campaign.

New Zealand – This is discussed in greater detail in the main text above. The quote here from Curtin (1989, p. 13) tells all, "The most unusual feature of military death in New Zealand over these five years was the fact that deaths from accident and battle exceeded deaths from disease … The high number of deaths in battle is evidence of heavy campaigning." In our minimal recoding, New Zealand is coded as campaign. In our extended recoding, Australia (for which data are derived from New Zealand) is also coded as campaign.

Senegal – Curtin's period is 1819-1838. During this period the French colony was basically just Gorée and St Louis islands in the mouth of the Senegal River. However, the French were very interested in expanding their commercial interests and started to build forts up the Senegal River at Dagna (1821) and Merinaghen (1822) (Oloruntimehin, 1974, p. 356). They also sent many missions into the interior. The French attempt to control trade started conflict.

"Thus, for instance, in 1832 the French in Senegal fought the Trarza Moors to establish their control over the gum trade. The same situation applied in the relation between the French, the Moors and the Jolof state of Walo in 1835. Military involvement of this nature was often protracted," (Oloruntimehin, 1974, pp. 356-367).

So once more it is incorrect that they were sitting in barracks. In our extended recoding, Senegal is recoded as campaign.

Trinidad and Tobago - this gets a mortality rate of 85 from the Windwards and Leewards 1817-1836. Curtin notes p. 25, "the central station was Barbados, but at times troops from the command served as far to the north as St Kitts and as far to the southeast as British Guiana".

This is significant. In 1823 was the massive Demerara Slave rebellion in Guyana. The beginning of this period also almost includes Bussa's Rebellion, a huge slave revolt in Barbados in 1816. A standard reference to this is Beckles (2006); see chapter 5 on Bussa's rebellion and aftermath. In 1817 they were still hanging people so there certainly was a large military force in operation and keeping the peace. The seminal book on the Demerara slave revolt is Da Costa (1994).

Blackburn notes (1988, p. 430) in the context of the repression of the Demerara rebellion, "The Governor called out well-armed troops and militia, including a detachment of one of the West India Regiment." Da Costa refers to this on page 217, so it appears likely that the troops stationed in Barbados saw action in both the Bussa and Demerara rebellions during this period.

In our extended recoding, we code Trinidad and Tobago as campaign.

South Africa. This rate comes from the Cape Colony 1818-1836. As far as we can find, Curtin says nothing specific about the presence or absence of military activity in Cape Colony. However, this period includes both the 5th and the 6th Xhosa Wars on the Eastern Frontier of the Cape. These involved British troops, etc. so we do not know exactly where the numbers in Curtin come from in terms of these campaigns – but the period clearly includes major campaigns.

The Xhosa Wars are discussed in all standard histories of South Africa, for example Thompson (2001, chapter 3).

In our extended recoding, South Africa is coded as campaign.

USA - this is for American troops 1829-1838. But US soldiers were obviously fighting Indian wars in this period. Again, Curtin does not discuss this number, but this period includes a number of Indian wars: the Second Seminole War in Florida, 1835-1842; The Black Hawk War 1832; and the Creek War of 1836.

Material on these wars appears in all standard histories of the US. For example, in the shorter Oxford History, Jones (1995, p. 118) writes, "The Seminole War of 1835-42 involved large-scale operations in the Florida swamps and cost the United States 1,500 men and $50 million."

This period also saw the forced removal of many Indians tribes following the passage of the 1830 Removal Act; see Banner (2005) – these removals were organized by the army.

In our extended recoding, the USA is recoded as campaign.

Summary

Our minimal recoding covers just Hong Kong and New Zealand. Our extended recoding covers those two countries, plus Jamaica, Malaysia, Singapore, Sri Lanka, Australia, Senegal, South Africa, Trinidad and Tobago, and the USA.

Additional References for the Appendices

Banner, Stuart (2005) *How the Indians Lost Their Land*, Cambridge: Belknap Press. Chapter 6.

Beckles, Hilary McD. (2006) *A History of Barbados*, New York Cambridge University Press.

Blackburn, Robin (1988) *The Overthrow of Colonial Slavery, 1776-1848*, London: Verso.

Chernozhukov, Victor and Christian Hansen. (2008) "The reduced form: A simple approach to inference with weak instruments." *Economics Letters*, 100(1): 68-71.

Da Costa, Emilia (1994) *Crowns of Glory, Tears of Blood: the Demerara Slave Rebellion of 1823*, New York: Oxford University Press.

Finlay, Keith and Leandro Magnusson (2009) "Implementing weak-instrument robust tests for a general class of instrumental-variables models." *Stata Journal*, 9(3): 398-421.

Jones, Maldwyn A. (1995) *The Limits of Liberty: American History 1607-1992*, 2nd edition, New York: Oxford University Press.

Mikusheva, Anna. (2010) "Robust confidence sets in the presence of weak instruments." *Journal of Econometrics*, 157(2): 236-247.

Mikusheva, Anna and Brian Poi (2006) "Tests and confidence sets with correct size in the simultaneous equations model with potentially weak instruments." *Stata Journal*, 6(3): 1-11.

Mills, L.A. (1966) *British Malaya 1824-1867*, New York: Oxford University Press.

Oloruntimehin, B.Olatunji (1974) "The Western Sudan and the Coming of the French, 1800-1893," in J.F.A. Ajayi and Michael Crowder eds. *The History of West Africa*, Volume 2, New York: Columbia University Press.

Peebles, Patrick (2006) *The History of Sri Lanka*, Westport Conn: Greenwood Press.

Reckord, Mary (1968) "The Jamaica Slave Rebellion of 1831," *Past and Present*, 39, 108-125.

Thompson, Leonard (2001) *A History of South Africa*, 3rd Ed. New Haven: Yale University Press.

Appendix Table 1A

First Stage Regressions, Alternative Samples

	Alternative samples							
	Original AJR series	Original AJR series, capped at 250	Original AJR series, capped at 150	Original AJR series, capped at 350	Original AJR series, capped at 280	Original AJR series, trimmed at 150	Original AJR series, trimmed at 250	Original AJR series, trimmed at 350
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dependent variable is protection against risk of expropriation							
No covariates	-0.61	-0.94	-1.12	-0.84	-0.89	-1.15	-1.16	-1.05
(standard error)	(0.15)	(0.16)	(0.18)	(0.16)	(0.16)	(0.21)	(0.19)	(0.18)
(clustered standard error)	(0.17)	(0.18)	(0.18)	(0.18)	(0.18)	(0.20)	(0.19)	(0.18)
Number of clusters	36	36	36	36	36	24	28	29
Number of observations	64	64	64	64	64	43	49	54
With latitude	-0.52	-0.86	-1.03	-0.75	-0.82	-1.05	-1.04	-0.95
(standard error)	(0.17)	(0.18)	(0.20)	(0.18)	(0.18)	(0.23)	(0.21)	(0.20)
(clustered standard error)	(0.19)	(0.20)	(0.20)	(0.20)	(0.20)	(0.22)	(0.21)	(0.20)
Number of clusters	36	36	36	36	36	24	28	29
Number of observations	64	64	64	64	64	43	49	54
Without neo-Europes	-0.4	-0.66	-0.8	-0.58	-0.63	-0.72	-0.84	-0.76
(standard error)	(0.15)	(0.18)	(0.22)	(0.17)	(0.18)	(0.27)	(0.24)	(0.21)
(clustered standard error)	(0.17)	(0.20)	(0.23)	(0.19)	(0.19)	(0.27)	(0.26)	(0.20)
Number of clusters	33	33	33	33	33	21	25	26
Number of observations	60	60	60	60	60	39	45	50
Without Africa	-1.21	-1.21	-1.23	-1.21	-1.21	-1.31	-1.21	-1.21
(standard error)	(0.20)	(0.20)	(0.20)	(0.20)	(0.20)	(0.21)	(0.20)	(0.20)
(clustered standard error)	(0.18)	(0.18)	(0.17)	(0.18)	(0.18)	(0.17)	(0.18)	(0.18)
Number of clusters	19	19	19	19	19	17	19	19
Number of observations	37	37	37	37	37	34	37	37
With continent dummies	-0.44	-0.81	-0.98	-0.7	-0.76	-1.17	-1.09	-0.92
(standard error)	(0.19)	(0.25)	(0.27)	(0.23)	(0.24)	(0.32)	(0.26)	(0.26)
(clustered standard error)	(0.20)	(0.25)	(0.27)	(0.24)	(0.24)	(0.31)	(0.26)	(0.26)
Number of clusters	36	36	36	36	36	24	28	29
Number of observations	64	64	64	64	64	43	49	54
With continent dummies and latitude	-0.35	-0.72	-0.89	-0.6	-0.67	-1.06	-0.95	-0.8
(standard error)	(0.20)	(0.26)	(0.29)	(0.24)	(0.25)	(0.32)	(0.27)	(0.26)
(clustered standard error)	(0.21)	(0.26)	(0.28)	(0.25)	(0.25)	(0.32)	(0.26)	(0.26)
Number of clusters	36	36	36	36	36	24	28	29
Number of observations	64	64	64	64	64	43	49	54
With percent of European descent in 1975	-0.42	-0.73	-0.88	-0.63	-0.69	-0.92	-0.96	-0.84
(standard error)	(0.17)	(0.19)	(0.20)	(0.18)	(0.18)	(0.22)	(0.20)	(0.20)
(clustered standard error)	(0.19)	(0.20)	(0.21)	(0.20)	(0.20)	(0.22)	(0.21)	(0.20)
Number of clusters	36	36	36	36	36	24	28	29
Number of observations	64	64	64	64	64	43	49	54
With malaria	-0.43	-0.81	-0.96	-0.71	-0.77	-1.07	-1	-0.88
(standard error)	(0.21)	(0.22)	(0.21)	(0.22)	(0.22)	(0.20)	(0.19)	(0.21)
(clustered standard error)	(0.24)	(0.24)	(0.22)	(0.25)	(0.24)	(0.21)	(0.20)	(0.23)
Number of clusters	35	35	35	35	35	23	27	28
Number of observations	62	62	62	62	62	41	47	52

OLS regressions, with one observation per country. Coefficients and standard errors for covariates, where included, are not reported to save space. Variables are from AJR (2001). Dependent variable is protection against risk of expropriation; independent variable is log settler mortality. Column 1 uses original settler mortality series from AJR (2001) as independent variable but includes Albouy's campaign dummy. Column 2 is the same as column 1 but caps mortality at 250 per 1,000 per annum (i.e., any mortality observation above 250 is set to equal 250). Column 3 is the same as column 1 but caps mortality at 150 per 1,000 per annum. Column 4 is the same as column 1 but caps mortality at 350. Column 5 is the same as column 1 but caps mortality at 280. Column 6 is the same as column 1 but trims mortality at 150 (i.e., any observation above 150 is dropped). Column 7 is the same as column 1 but trims mortality at 250. Column 8 is the same as column 1 but trims mortality at 350.

Appendix Table 1B

Second Stage Regressions, Alternative Samples

	Alternative samples							
	Original AJR series	Original AJR series, capped at 250	Original AJR series, capped at 150	Original AJR series, capped at 350	Original AJR series, capped at 280	Original AJR series, trimmed at 150	Original AJR series, trimmed at 250	Original AJR series, trimmed at 350
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dependent variable is log GDP per capita in 1995							
No covariates	0.93	0.82	0.78	0.85	0.83	0.66	0.64	0.65
AR confidence set	[0.69,1.40]	[0.62,1.14]	[0.58,1.08]	[0.64,1.21]	[0.62,1.16]	[0.43,0.96]	[0.44,0.89]	[0.46,0.89]
AR confidence set, clustered	[0.67,1.72]	[0.61,1.19]	[0.58,1.09]	[0.64,1.31]	[0.62,1.23]	[0.42,0.98]	[0.45,0.91]	[0.49,0.87]
F-stat, first stage	23.34	35.55	37.53	31.93	34.19	27.53	37.35	37.08
F-stat, first stage, clustered	12.45	28.09	37.09	21.9	25.87	32.56	36.78	33.04
With latitude	0.96	0.79	0.74	0.85	0.81	0.62	0.60	0.60
AR confidence set	[0.65,1.78]	[0.55,1.24]	[0.51,1.14]	[0.58,1.37]	[0.56,1.28]	[0.35,0.99]	[0.35,0.93]	[0.36,0.93]
AR confidence set, clustered	[0.65,2.45]	[0.57,1.18]	[0.53,1.03]	[0.61,1.40]	[0.57,1.24]	[0.36,0.91]	[0.36,0.85]	[0.38,0.81]
F-stat, first stage	13.48	21.82	23.58	19.15	20.79	19.92	23.52	22.01
F-stat, first stage, clustered	7.3	19.26	25.57	14.24	17.49	22.16	24.54	23.53
Without neo-Europes	1.24	1.04	1.00	1.11	1.06	0.85	0.74	0.75
AR confidence set	[0.78,3.09]	[0.68,1.99]	[0.64,1.93]	[0.72,2.24]	[0.69,2.06]	[0.34,2.42]	[0.39,1.40]	[0.43,1.33]
AR confidence set, clustered	[0.76,5.43]	[0.65,2.10]	[0.62,1.89]	[0.70,2.56]	[0.67,2.21]	[0.24,2.14]	[0.35,1.53]	[0.44,1.26]
F-stat, first stage	8.89	13.22	12.69	11.92	12.78	6.89	13.07	13.95
F-stat, first stage, clustered	5.54	11.27	12.56	9.21	10.73	7.2	10.56	14
Without Africa	0.61	0.61	0.60	0.61	0.61	0.58	0.61	0.61
AR confidence set	[0.41,0.87]	[0.41,0.87]	[0.41,0.86]	[0.41,0.87]	[0.41,0.87]	[0.39,0.80]	[0.41,0.87]	[0.41,0.87]
AR confidence set, clustered	[0.45,0.85]	[0.45,0.85]	[0.44,0.83]	[0.45,0.85]	[0.45,0.85]	[0.45,0.74]	[0.45,0.85]	[0.45,0.85]
F-stat, first stage	30.62	30.62	31.46	30.62	30.62	31.03	30.62	30.62
F-stat, first stage, clustered	45.98	45.98	50.78	45.98	45.98	63.09	45.98	45.98
With continent dummies	0.97	0.78	0.78	0.84	0.80	0.75	0.68	0.65
AR confidence set	[0.59,3.20]	[0.52,1.42]	[0.53,1.34]	[0.54,1.67]	[0.52,1.49]	[0.49,1.23]	[0.46,1.07]	[0.41,1.05]
AR confidence set, clustered	[0.52,4.87]	[0.45,1.43]	[0.47,1.33]	[0.47,1.70]	[0.45,1.49]	[0.47,1.30]	[0.41,1.14]	[0.34,1.05]
F-stat, first stage	6.5	13.32	15.53	11	12.42	16.67	21.07	16.64
F-stat, first stage, clustered	4.68	10.61	13.29	8.62	9.97	13.84	17.69	12.78
With continent dummies and latitude	1.07	0.80	0.80	0.88	0.82	0.74	0.69	0.66
AR confidence set	[-∞,-27.22] U [0.57,∞]	[0.48,1.93]	[0.49,1.66]	[0.51,2.74]	[0.48,2.13]	[0.45,1.37]	[0.41,1.28]	[0.35,1.30]
AR confidence set, clustered	[-∞,-4.72] U [0.44,∞]	[0.30,1.53]	[0.39,1.35]	[0.34,2.17]	[0.30,1.65]	[0.40,1.24]	[0.29,1.16]	[0.09,1.09]
F-stat, first stage	3.71	8.52	10.52	6.73	7.8	12.67	13.4	10.07
F-stat, first stage, clustered	2.72	7.74	10.22	5.87	7.1	11.28	12.84	9.64
With percent of European descent in 1975	0.92	0.71	0.65	0.77	0.73	0.54	0.51	0.51
AR confidence set	[0.55,2.31]	[0.44,1.27]	[0.39,1.12]	[0.48,1.48]	[0.45,1.33]	[0.23,0.96]	[0.26,0.85]	[0.26,0.85]
AR confidence set, clustered	[0.53,4.31]	[0.36,1.21]	[0.30,1.04]	[0.43,1.49]	[0.39,1.27]	[0.10,0.93]	[0.16,0.84]	[0.14,0.80]
F-stat, first stage	8.67	15.32	16.78	13.12	14.48	15.37	20.64	18.55
F-stat, first stage, clustered	4.92	12.92	17.56	9.65	11.76	17.7	20.46	17.26
With malaria	0.67	0.52	0.54	0.55	0.52	0.54	0.51	0.47
AR confidence set	[0.29,2.93]	[0.27,0.95]	[0.31,0.91]	[0.27,1.11]	[0.26,0.98]	[0.32,0.86]	[0.28,0.83]	[0.21,0.80]
AR confidence set, clustered	[-∞,-3.76] U [0.25,∞]	[0.23,0.89]	[0.32,0.81]	[0.23,1.12]	[0.21,0.94]	[0.37,0.75]	[0.29,0.76]	[0.18,0.74]
F-stat, first stage	5.38	13.95	17.98	10.92	12.73	23.45	25.35	17.62
F-stat, first stage, clustered	3.11	11.45	19.03	7.9	9.99	26.02	25.88	14.63

2SLS regressions, with one observation per country, corresponding to first-stage regressions in Appendix Table 1A. Variables are from AJR (2001). Dependent variable is log GDP per capita in 1995. Right-hand side variable is protection against expropriation, instrumented by log settler mortality. Column 1 uses original settler mortality series from AJR (2001) as the instrument. Column 2 is the same as column 1 but caps mortality at 250 per 1,000 per annum (i.e., any mortality observation above 250 is set to equal 250). Column 3 is the same as column 1 but caps mortality at 150 per 1,000 per annum. Column 4 is the same as column 1 but caps mortality at 350. Column 5 is the same as column 1 but caps mortality at 280. Column 6 is the same as column 1 but trims mortality at 150 (i.e., any observation above 150 is dropped). Column 7 is the same as column 1 but trims mortality at 250. Column 8 is the same as column 1 but trims mortality at 350.