Population and Civil War*

Daron Acemoglu
MIT and NBER

Leopoldo Ferguson
Universidad de los Andes

Simon Johnson
MIT and NBER

December 2016

Abstract

Medical and public health innovations in the 1940s quickly resulted in significant health improvements around the world. Countries with initially higher mortality from infectious diseases experienced greater increases in life expectancy, population, and – over the following 40 years – social conflict. This result is robust across alternative measures of conflict and is not driven by differential trends between countries with varying baseline characteristics. At least during this time period, a faster increase in population made social conflict more likely, probably because it increased competition for scarce resources in low income countries.

Keywords: civil war, population growth
JEL Classification: J1, 011, O15, Q56

*We thank seminar participants at MIT's development lunch, the University of Chicago, the XXII Annual Conference of the European Society for Population Economics at UCL, Universidad de los Andes, and Universidad del Rosario. Ioannis Tokatlidis provided superb research assistance in the early stages of this project.

1 Introduction

Civil war is a complex phenomenon with multiple causes. However, understanding potential common drivers is important given the enormous human suffering caused by violence.[1] This paper digs into one potential cause – population pressure at the national level.

The world's population is forecast to rise from its current level of around 7.35 billion to over 11.2 billion by 2100. Today's "more developed regions", as classified by the UN, had population of 1.25 billion in 2015 and are expected to have roughly the same number of inhabitants – 1.28 billion – in 2100. In contrast, "least developed countries" had population of 954 million in 2010 and are projected to reach over 3.1 billion in 2100. The UN expects that the population of Africa will rise from its current level of just over 1.1 billion to over 4.3 billion within a century.[2] Seen over the past two centuries – and in fact over all of human history – the century of population increase from 1950 to 2050 stands out dramatically.

A further major global push to improve public health in low income countries is now under way. The coming improvement in the lives of millions may go down as one of the great successes in human history.[3] However, we should consider potentially more complicated consequences of this major development. Should we expect higher population levels to put greater pressure on limited resources, such as land, and result in more social conflict – including, at least potentially, a greater risk of social conflict? Some relevant recent historical experience – the topic of this paper – suggests grounds for concern.

Beginning in the 1940s, there was an *international epidemiological transition* driven by the introduction of new chemicals, drugs, and public health measures (Acemoglu and Johnson

[1]See Abadie and Gardeazabal (2003), Fearon and Laitin (2003), Blattman and Miguel (2010), and Fergusson, Ibáñez and Riaño (2015).

[2]The data in this paragraph are from the UN's 2015 long-term population projections using the medium fertility scenario (United Nations, 2015). Mid-year world population in 2015 was estimated as 7.349 billion, forecast to rise to 9.725 billion by 2050 and 11.213 billion by 2100. The population of Africa was, in these estimates, 1.186 billion in 2015; it will be 2.477 billion in 2050 and 4.387 billion in 2100 (of whom 3.934 billion will be in sub-Saharan Africa). Most of the potential forecast error around these estimates is due to uncertainty about what will happen in low income countries. In the UN's high fertility scenario, world population reaches 16.6 billion in 2100, of which nearly 1.9 billion are in today's high income countries and 14.7 billion are in today's low income countries. In the low fertility scenario, world population is 7.2 billion in 2100, of which 842 million are in today's rich countries and just under 6.5 billion in today's poor countries. Even in the low fertility scenario, the population of sub-Saharan Africa rises from 962 million in 2015 to 2.75 billion in 2100.

[3]For example, in laying out their most recent goals for global health, philanthropists Bill and Melinda Gates (2015) argue, "In 1990, one in ten children in the world died before age 5. Today, it's one in 20. By 2030, that number will be one in 40. Almost all countries will include vaccines for diarrhea and pneumonia, two of the biggest killers of children, in their immunization programs. Better sanitation — through simple actions like hand-washing as well as innovations like new toilets designed especially for poor places — will cut the spread of disease dramatically" (p. 5) And, "In 15 years, we'll be poised to send malaria the way of smallpox and polio" (p.8). The worldwide under-5 mortality rate fell from 9% in 1990 to 4.6% in the latest data; they predict it will fall to 2.3% by 2030.

2007). The effect on life expectancy – and on population – varied by country, depending in part on the extent to which inhabitants were previously affected by infectious disease. Some initially low income countries quickly experienced large increases in life expectancy. Birth rates fell, but not as quickly as mortality declined, with the result that population increased in some places at rates that were high by historical standards.

Figure 1 provides some global context – the annualized average growth rate of world population over the past 12,000 years and looking forward to 2100.[4] World population growth rate exceeded 0.5 percent per annum for the first time in the period 1800-1850, mostly due to the increase in size of some European countries. For example, the number of people living in the British Isles, during its most rapid century of population growth (1800-1900) increased from 16 million to 42 million – a remarkable increase by any previous historical standard. Other European countries experienced population increases and a demographic transition – from high mortality to low mortality – along similar lines.[5]

Most countries outside of Europe did not experience the same scale of health improvements during the nineteenth and early twentieth century. However, this situation changed dramatically after 1940, partly as a result of new drugs and practices introduced after this date, such as DDT, penicillin and streptomycin. Low income countries with initially high mortality experienced rapid health improvements and longer life expectancy – and the growth rate of global population increased.

Sri Lanka illustrates population dynamics in initially poor countries with high mortality. Malaria and other infectious diseases previously constituted a major burden but were substantially addressed in the 1940s and early 1950s. Birth rates fell, but not as quickly as mortality declined. The country's population consequently grew at 2.8 percent per annum from 1950 to 1975, slowing to 1.13 percent per annum through 2010. Over the next 40 years, from 2010 to 2050, the population of Sri Lanka is likely to rise at a rate of just 0.34 percent per annum.[6] From 1950 to 2050, Sri Lanka's population is expected to increase from around 7 million to just over 23 million. If these estimates are correct, the total increase over a century will prove higher than the cumulative British experience during the 19th century. The main difference

[4] The historical data through 1950 are from McEvedy and Jones (1978). For 2000 and 2050 we use estimates from the Census Bureau and for 2100 we use the UN 2010 revision population projections.

[5] The major differences were in so-called settler colonies, such as the USA, Canada, Australia, New Zealand, and Argentina. In those cases natural rates of population increase were supplemented by high rates of immigration. Of course, indigenous populations typically declined over the period of European settlement.

[6] These data are from McEvedy and Jones (1978) through 1975 and then from the UN population projections (2010 revision).

from the earlier European experience is that, at its peak rate, Sri Lanka's population grew faster and the total increase was relatively compressed in time.[7] This was not an isolated experience – in many low income countries after 1940, there was a population growth shock.

In some of the world's poorest countries today, the potential population increase over the next 40 years could easily match or even exceed the most dramatic developing country population shocks of the post-1940 period.[8] For example, the population of Nigeria was 30 million in 1950 and 160 million in 2010; it could reach 440 million by 2050 and close to 1 billion by 2100.[9] The potential increase in other other sub-Saharan countries is similar. The rate of population increase may now be slowing, but an increase of more than five-fold over a century (1950-2050) seems likely and a ten-fold increase is entirely possible.[10] The expected increases in parts of North Africa and the Middle East are smaller, but the population of those regions could well double between 2010 and 2100.[11]

The rate of population increase in some of the most troubled countries in the Middle East is striking. For example, Syria had a population of around 1.75 million in 1900 and 3.25 million in 1950. Syria's population had risen to 18.5 million in 2015 and – before the civil war – was projected to rise to over 35 million by 2050 and 38 million by 2100. Just as an indication of population pressure, over the century from 1950 to 2050, population was expected to increase by nearly ten-fold – although now, of course, massive emigration is likely to change this picture.[12]

Motivated by what may happen over the next 50 or 100 years, we focus in this paper on

[7] In England, Wales, Scotland, and Ireland, from 1800 to 1850, the annualized rate of population growth was 1.11 percent. Population growth slowed markedly in the 20th century – to 0.5 percent per annum from 1900 to 1950, and to 0.35 percent per annum in the 25 years to 1975.

[8] There is also an important difference between today and 1940, at the level of the world as a whole. The rate of world population growth has definitely slowed – down from a peak of over 2 percent per annum, in the late 1980s – and is unlikely now to accelerate, in part because the most populous countries in the world (including China and India) now have relatively low fertility rates. This is reflected in Figure 1.

[9] There is considerable uncertainty regarding the speed of fertility decline in today's low income and relatively high mortality countries. Still under all plausible scenarios, the population increases will be large. For example, in the UN's low fertility scenario, the population of Nigeria rises from 160 million in 2010 to 644 million in 2100, which would be one of the largest increases in population over a century that the world has ever seen. In the high fertility scenario Nigeria has a population of almost 1.3 billion in 2100.

[10] Contrast these increases with the increase in US population following the much discussed "baby boom" after 1945. In the US, population increased "only" around 70% from 1940 to 1980 – starting at a level of about 130 million in 1940 and reaching roughly 228 million in 1980.

[11] In the UN's medium fertility scenario, population in Northern Africa (the Mediterranean littoral, plus Western Sahara and Sudan) rises from 200 million in 2010 to 370 million in 2100. The population of Western Asia (the Middle East, Turkey, Cyrpus, and the Caucasus) rises from 232 million in 2010 to just over 400 million in 2100.

[12] The 1900 and 1950 data are from McEvedy and Jones. They report population for Syria and Lebanon from 400BC to 1975. This population was 2.25m in 1AD and fluctuated between 1.5m and 1.75m in the following 18 1/2 centuries.

Iraq is another country in the midst of a remarkable population increase. Population in 2015 was 36.4 million. In 2050, population is predicted to be 83.7 million.

an historical episode with potential relevance: between 1940 (when global health technology improved dramatically) and 1980 (just before HIV-AIDS spread as a global disease). We instrument for population growth based on the initial (1940) distribution of mortality from various diseases around the world and the dates of global "interventions" that brought down mortality from those diseases. Most of the medical and public health breakthroughs in this period originated in a few industrialized countries and can reasonably be seen as exogenous to development prospects in the rest of the world. Our instrument also does not depend in any way on when a particular country adopted better public health measures or how effectively these measures were applied.[13]

We control for other potential determinants of civil war both directly (either with a specific measure drawn from the literature or using country level fixed effects) and through allowing time varying fixed effects (allowing differential trends based on initial conditions). In this regression framework, countries with higher exogenous increases in population experienced more social conflict in the post-1940 period. Across alternative definitions of civil war and social conflict, instrumented changes in population have a robust significant positive effect on the share of years per decade in which a country experienced civil war or other forms of violent social conflict.

Our methodology cannot distinguish whether it is the level of population or unusually large changes that cause violence. More importantly, our theoretical framework suggests the key Malthusian mechanisms are not about increases in population per se. Instead, the expectation is that violence is more likely both when there is a larger population for a given level of total productivity, or with increases in population that are unusually large relative to the technological and other processes that tend to increase total factor productivity.

The magnitude of our estimates indicate that the effect is large. A rise in log population of about 0.68 from 1940 to 1980, corresponding to the average change in log population in our sample of countries, causes roughly 4.2 additional years of full-blown civil war in the 1980s relative to the 1940s (or 1950s). When considering lower intensity conflicts, the corresponding effect is similar – about 3.9 more years in conflict in 1980 as a result of the increase in population from 1940.[14]

[13]This analysis is relevant to understanding the population pressures that may now develop around the world, but because our study requires detailed death-by-disease data, our base sample contains some but by no means all countries that were low income in 1940. In particular, we do not have data on many sub-Saharan countries, and we should be cautious about the extent to which these places may be on a different trajectory.

[14] Our sample contains 59 non-Eastern Europe countries (16 countries in Asia, 17 in Europe, 17 in Latin America, five in Africa, Australia, New Zealand, Canada, and the US), and 6 Eastern European countries. Of

The 1940s was of course a decade of global war. To take this into account we run panel regressions that exploit decade-by-decade changes in population from 1940 to 1980. We also verify that our results hold when excluding the countries demographically most affected by World War II. As an additional robustness check, we assign the level of conflict of the 1950s to the 1940s. Finally, we run specifications ignoring the World War II years. In all cases, we find similar results. We also create a new definition of civil war, based on a relative threshold of violence. Our results are substantially robust across this set of specification checks.

Section 2 reviews the literature relevant to the question of whether population increases can result in more social conflict. In section 3 we present a simple motivating theory capturing Malthusian mechanisms that may lead from population to conflict. Section 4 describes our data, and Section 4.3 presents ordinary least square (OLS) results. Section 5 discusses our identification strategy, and Section 6 shows our main results from two-stage least squares (2SLS) estimates. Section 7 presents a series of robustness checks on our estimates. Section 8 concludes.

2 Literature On Population And Conflict

This paper is related to several other strands of research. An important line of research uses cross-country evidence to study the causes of civil war. Following contributions such as those of Collier and Hoeffler (1998; 2004) and Fearon and Laitin (2003), scholars have emphasized poverty, inequality, weak institutions, political grievances, and ethnic divisions as explanations for the outbreak and persistence of civil war. With a few notable exceptions, however, this literature does not fully address the possibility that reverse causality, or omitted variables bias, drives the observed correlations.[15] Indeed, Blattman and Miguel (2010) conclude in their survey of the literature that "further cross-country regressions will only be useful if they distinguish between competing explanations using more credible econometric methods for establishing causality" (p. 8).

---

these 65 countries, 13 countries lack population data and/or had not yet been created in 1940 (five in Africa, seven in Asia, and Russia). Also, Austria is excluded in 1940 when the dependent variables are from COW since it enters the COW state system in the 1950s.

Our sample has a good coverage of most regions in the world, with the exception of Africa. Given the incidence of violent civil conflict in Africa, this is an important limitation, but one that we cannot overcome with the available data. In particular, most sub-Saharan African countries lack reliable data on causes of death disaggregated by disease dating back to the 1940s, and this is essential for our identification strategy.

[15] Exceptions include Miguel, Shanker Satyanath and Sergenti (2004), who use annual rainfall growth as an instrument for income growth in sub-Saharan Africa; Besley and Persson (2008), who rely on plausibly exogenous international commodity price movements; and Brückner (2010), which we review below.

More broadly, population has not been a prime focus in the literature on the "economics of conflict" (see, for example, the survey by Garfinkel and Skaperdas, 2007).[16] However, there has been a lively debate on the effects of population pressure on violent conflict in other disciplines, including political science[17]. There is heated debate between people who highlight the importance of "Malthusian" channels that lead to conflict and those who remain more skeptical about the practical relevance of such issues. The evidence so far has not been conclusive.[18]

For example, Homer-Dixon (1991, 1999) studies the connection between population growth, pressure on environmental resources, and conflict – finding that poor countries are in general more vulnerable to environmentally-induced conflicts. Moreover, while recognizing (as do many anti-Malthusians) resilience and adaptability in human-environmental systems, he contends that "as population grows and environmental damage progresses, policymakers will have less and less capacity to intervene to keep this damage from producing serious social disruption, including conflict" (Homer-Dixon, 1991). During the 1990s, Homer-Dixon and his collaborators on the Project on Environment, Population and Security (EPS), often referred to as the Toronto Group, produced a number of case studies on "environmental conflict" which, though not concerned solely with population growth, did underscore its importance as potential source of environmental scarcity and consequently conflict (Homer-Dixon, 1994).

The overall approach of these studies on environmental security has come under attack from a number of researchers. Richards (1996) dismisses what he calls the "New Barbarism" theory of conflict put forward in Kaplan's (1994) essay. In particular, for the case of Sierra

[16] In most of the empirical economics literature on conflict, population is a control variable (often with a positive sign), but it is rarely the prime focus and there is no attempt to control for its endogeneity. For instance, in Sambanis (2002) review of this research, the role of population is hardly mentioned. Collier and Hoeffler (2004) report a positive coefficient on population, which the authors interpret as consistent with either a greed or grievance story for conflict, but their regressions for a panel of countries do not control for country fixed-effects and thus may well be driven by omitted country-specific characteristics. In Fearon and Laitin's (2003) study of conflict onset, the positive coefficient on population disappears once fixed effects are included in the regression. Miguel, Satyanath, and Sergenti (2004) also report a positive coefficient, but their focus is on the effect of income on conflict.

[17] The connection between population and conflict has also received significant public attention, as testified by Robert Kaplan's famous 1994 essay "The Coming Anarchy," in turn heavily influenced by Homer-Dixon.

[18] Tir and Diehl (1998) examine the Correlates of War dataset to evaluate the impact of population growth and density on international conflict involvement, initiation, and escalation over the period 1930-89. They find that population growth pressures have a significant impact on military conflict involvement, especially in poor countries, but no correlation with conflict initiation or escalation, or between population density and conflict. Hauge and Ellingsen (1998) find that factors like deforestation, land degradation, and scarce supply of freshwater, alone and in combination with high population density, increase the risk of domestic armed conflict, especially low-level conflict, in the period 1980-92. However, economic and political variables prove more decisive than environmental scarcity in predicting the incidence of domestic armed conflict. For more studies along these lines, see the special 1998 issue of the Journal of Peace Research and Diehl and Gleditsch (2001).

Leone, he pushes back against the neo-Malthusians who "believe that ordinary Africans will 'sink their ship' by over-production, unless checked by famine, war and disease [and that] war is a process through which the poor in Africa will learn its limits" (p. 121). Instead, he notes that the process of forest conversion in Sierra Leone has taken place over many centuries, and that local land-users have responded in sensible way to its different phases, with no evidence of environmental degradation spiralling out of control prior to or during the years of civil war. He concludes that in Sierra Leone "war is a consequence of political collapse and state recession, not environmental pressure (...) Violence has been incubated in forest fastnesses. [The problem] is too much forest, not too little" (p. 124).[19]

Overall, it is fair to say that the evidence to date on the impact of population increases has not been conclusive, in part because most studies have not adequately addressed the potential endogeneity of population in regressions with social conflict as the dependent variable. Brückner (2010) is the lone exception – a paper that explicitly tackles causality issues in a study of population size and civil conflict. Brückner studies a panel of 37 Sub-Saharan countries during the period 1981–2004 and uses randomly occurring droughts as an instrumental variable for population. He finds that instrumented population size has an economically meaningful and statistically significant effect on African civil conflict. The most important threat to the validity of his estimates is that droughts may affect conflict through their effect on other variables, such as income, and not via population.[20]

The empirical question addressed in this paper is relevant for a longer tradition of theories dating back at least to Malthus, who suggested that higher population density may lead to more social conflict. Along these lines, we use a simple theoretical framework in which larger populations, absent a corresponding increase in resources and technology, exacerbate the competition for resources and increase the likelihood of civil wars. Our paper contributes a new way to address the potential endogeneity of population in regressions for conflict – and our

[19] Gleditsch (1998) argues that the environmental security line of research fails to qualify as "systematic" quantitative or comparative research, and violates the rules of quasi-experimental methodology.

His full critique includes 9 more specific "problems" of the literature: 1. There is a lack of clarity over what is meant by "environmental conflict"; 2. Researchers engage in definitional and polemical exercises rather than analysis; 3. Important variables are neglected, notably political and economic factors, which have a strong influence on conflict and mediate the influence of resources and environmental factors; 4. Some models become so large and complex that they are virtually untestable; 5. Cases are selected on values of the dependent variable; 6. The causality of the relationship is reversed; 7. Postulated events in the future are cited as empirical evidence; 8. Studies fail to distinguish between foreign and domestic conflict; and, 9. Confusion reigns about the appropriate level of analysis.

[20] Brückner (2010) is aware of this issue and estimates the impact of population on civil conflict risk conditional on per capita GDP. He also uses rainfall and international commodity prices as additional instrumental variables to deal with the endogeneity of income.

results suggest that there should now be cause for concern.

To offset this greater tendency towards conflict, countries either have to have enough growth or find ways to mitigate conflict. Nothing in our results should be construed as opposing health improvements or population growth. But it is important to put the post-1940 history – and our likely global history – in a broader perspective.

The effect of population growth on civil conflict can help explain a puzzling fact in the literature on long-run growth after 1940, documented in Figure 2. The international epidemiological transition produced large increases in population, especially for initially poor countries (2a and 2b), and significant convergence in health conditions around the world (2c). By the year 2000, the gap in average life expectancy at birth between initially rich and initially poor countries was reduced to about a half of its 1930 level, measured in absolute terms. However, in spite of an extensive microeconomic literature showing that improving health can improve individual economic outcomes and potentially accelerate economic growth, no such convergence is apparent when examining output per capita (2d). While average log GDP per capita for initially poor, middle-income, and rich countries has trended upwards since the 1930s, poorer countries have not been able to outgrow and catch-up with richer countries.

It remains to be seen if this form of economic convergence will be stronger over the next half century, but Figure 3 suggests increased social conflict may be one reason behind the lack of convergence to date. Since 1940, conflict incidence increased especially in poor countries, which experienced the largest increases in life expectancy and population. This is clear whether we measure the fraction of the decade with internal conflict using each of our alternative data sources (COW in 3a, UCDP/PRIO in 3b, or Fearon and Laitin in 3c) or if we look at the (log of) total deaths per yeas (3d).

Pinker (2011) argues that violence and social conflict of all forms are in secular decline. This is grounds for optimism and perhaps this global trend will predominate even in countries that are now experiencing rapid population growth. Hopefully, we have put civil conflict largely behind us in all corners of the globe. But it may be unwise to bet too much on that proposition. In terms of population pressure, experience in the 1940-1980 period is repeated today in countries such as Iraq, Syria, Afghanistan, Libya, and Myanmar.[21] People living

[21] The population of Iraq was 5.25 million in 1950, 11 million in 1975, 31 million in 2010, and is likely to reach 71 million in 2050 and 106 million in 2100.

Syria had a population of 3.25 million in 1950 and 7.25 million in 1975. This population is projected to rise from 21.5 million in 2010 to 36.7 million in 2050 and 40 million in 2100.

Afghanistan's population was 9 million in 1950 and 16 million in 1975; these totals will likely rise from 28.4 million in 2010 to 56.5 million in 2050 and 59 million in 2100.

today in low income countries, both at the lower end of the income scale and in the emerging middle class, need jobs, higher incomes, and better living standards, along with improved health.

3 Malthusian Mechanisms

In this section, we present a simple framework capturing the Malthusian idea that population growth may lead to social conflict. The basic idea is that higher population generates greater rents for a fixed factor relative to labor, and this form of scarcity makes conflict more likely. For less-developed economies in 1940 or today, it makes sense to think of land as the scarce factor.

One point of this framework is to emphasize that population growth does not necessarily lead to conflict. Indeed, it is not necessarily true with constant returns to scale to variable factors. However, we show that when greater population increases scarcity, it also makes conflict more likely.

Suppose that aggregate output is given by a constant returns to scale production function with a fixed factor, Z, labor, N, and technology, A:

$$Y = F(Z, N, A) \equiv f(N), \tag{1}$$

where $F(\cdot)$ exhibits constant returns to scale in (Z, N) and f gives output as a function of labor, holding technology and Z constant. Thus, if N increases with A constant, output per worker, $f(N)/N$, declines. However, if increases in labor – which we use as a synonym for population – are accompanied by increases in the technology parameter A, output per worker can remain constant, thus avoiding scarcity.

We assume the following simple allocation of resources. Each individual i in society supplies one unit of labor inelastically and also owns a fraction θ_i of land. For simplicity, we also suppose markets are competitive, though this is not important for our analysis. With these assumptions, individual income and consumption is given by

$$c_i(N, \theta_i) = f'(N) + \theta_i \left[f(N) - N f'(N) \right]. \tag{2}$$

Libya's population was 1 million in 1950 and 2.5 million in 1975, and 6 million in 2010. The number of inhabitants is projected to reach 8.35 million in 2050 and 7.6 million in 2100.

The population in Myanmar was 19 million in 1950 and 30 million in 1975. These totals are forecast to rise from 52 million in 2010 to 58.6 million in 2050 and then fall back to 47 million in 2100.

(These data are from McEvedy and Jones, 1978, for 1950 and 1975, and the UN 2012 population projections, medium variant, for 2010, 2050, and 2100.)

The key observation from equation (2) is that the marginal increase in an individual's consumption from an increase in his landholdings is larger when population increases,

$$\frac{\partial^2 c_i}{\partial N \partial \theta_i} = -N f''(N) > 0.$$

Land shares matter more for consumption when population is larger. The intuition is simple: with higher N, land rents are more important relative to wages due to diminishing returns. This implies a *Malthusian channel* to conflict when control over land can be contested with violence.

To explore this channel, imagine the society consists of two groups, 1 and 2. All members within a group are identical. To simplify the discussion we suppose both groups are of size $N/2$ and population growth leaves relative shares unchanged. To capture the disruption costs of conflict, assume that if a group initiates conflict, then this reduces total output to a fraction $(1-\rho)$ of what it would have been without conflict.

Group j has probability p_j of winning the conflict and if it does win, it captures a fraction λ_{-j} of the land of the other group, where λ is loosely an inverse measure of the "specificity of assets" to groups (or to individuals within a group). With probability $p_{-j} = 1 - p_j$, group j loses the conflict and a fraction λ_j of its land. Also for simplicity, any advantage of being the first mover is ignored and there are no deaths from any conflict. Also, as discussed below, voluntary concessions to avoid civil war are ignored. Finally, assume that all agents are risk neutral. Then, the expected benefits to conflict, $\pi_j(N, \theta, \lambda, \rho)$, for group j is given by,

$$\begin{aligned} \pi_j(N,\theta,\lambda,\rho) &= -\rho\left\{F'(N) + \theta_j\left[f(N) - Nf'(N)\right]\right\} \\ &\quad + (1-\rho)\left[p_j\lambda_{-j}\theta_{-j} - p_{-j}\lambda_j\theta_j\right]\left[f(N) - Nf'(N)\right]. \end{aligned} \tag{3}$$

In (3),the first line captures the deadweight destructive costs of conflict. The second line captures potential benefits, amounting to the undestroyed expected additional land rents that will be expropriated with violence. For there to exist equilibrium conflict, a necessary (but not sufficient condition) is for:

$$p_j\lambda_{-j}\theta_{-j} - p_{-j}\lambda_j\theta_j \neq 0.$$

If this holds, one of the groups will have potential gains from conflict—e.g., group j. These gains are likely to be higher when θ_{-j} and λ_{-j} are high, so greater inequality of resources between the two groups and a lower degree of specificity will contribute to the potential gains from conflict. But even in this case $\pi_j(N,\theta,\lambda,\rho) < 0$ is possible for both groups because of the first term in (3)—*cost of disruption.*

The same reasoning as in our discussion of equation (2) implies that whenever $\pi_j(N,\theta,\lambda,\rho) = 0$,

$$\frac{\partial \pi_j(N,\theta,\lambda,\rho)}{\partial N} > 0.$$

Therefore, an increase in population makes the group that is more likely to initiate civil war more "pro civil war." As noted before, this result does not apply when N increases in tandem with A. This observation is important, in the sense that the Malthusian mechanism says nothing about increases in population per se. Rather, the predictions are about the level of population for given A or for increases in population that are unusually large relative to the technological and other processes that tend to increase A.

This simple framework generates other intuitive comparative static results. Greater inequality (θ) between groups makes conflict more likely. Lower disruption costs (lower ρ) and lower asset specificity (higher λ), makes conflict more likely. The point about asset specificity is linked to the importance of natural resources and agriculture relative to human capital and industry. In particular, modern capitalism depends on production processes – such as factories and long supply chains – that can be easily disrupted with violence. When traditional production methods are prevalent, for instance when the main form of capital is land, the costs of violence are relatively smaller. Presumably, the productivity of land is harder to destroy than the productivity of a factory. Also, human capital is hard to expropriate through violence and, unlike land, can move to other regions or countries when there is an outbreak of violence (Acemoglu and Robinson, 2006).

Finally, a central question that we have ignored is why is conflict not prevented by more efficient ways of redistributing resources. A plausible explanation concerns commitment problems (Acemoglu and Robinson, 2001, 2006; Fearon, 1998, 2004; Powell, 2006; Acemoglu, Egorov and Sonin, 2008). To see this, consider the same environment in a dynamic setting, but in each period there is a probability $q < 1$ that either group can initiate civil war. Assume all agents have discount factor $\beta \in (0,1)$. To simplify the discussion, assume that, after civil war, there is a permanent redistribution of resources and never any social conflict again. Also to simplify, only cash transfers are feasible and group 1 is the one considering civil war.

In this context, the benefits from civil war for group 1 are proportional to $1/(1-\beta)$ because of discounting. If the group is sufficiently patient (β is high enough), then cash transfers in a given period are not sufficient to offset this gain. But group 2 cannot make a credible promise to make the cash transfers in the future once the window of opportunity for civil war disappears.

In this setting, civil wars arise along the equilibrium path even though more efficient ways of dealing with conflict exist. In particular, fix $\beta \in (0, 1)$, then there exists $\bar{q}$ such that for all $q < \bar{q}$, the Markov Perfect Equilibrium will involve equilibrium civil war. Also, there exists $\hat{q} < \bar{q}$, so that for all $q < \hat{q}$, all Subgame Perfect Equilibria involve civil war.

4 Data[22]

In our baseline analysis, we measure conflict as the ratio of number of years in conflict to total years for a period around a reference date t (where, typically, $t = 1940, 1950, ..., 1980$) and the conflict occurs in the decade that followed that date. This measure captures conflict incidence, rather than the precise timing of a conflict – this is appealing because we are interested in a relatively long-term phenomenon: increases in population over a period of several decades, and the potential response in terms of greater social conflict. Relatedly, datasets sometimes disagree on the exact year when a conflict began, but there are typically fewer differences regarding the incidence of conflict within a decade.

Our baseline dataset is version 3 of the Correlates of War (henceforth COW) dataset (Sarkees and Schafer, 2000). In these data, a civil war is defined as a war fought within state borders, between government and non-government forces, where the central government is actively involved in military action, with effective resistance for both sides, and with at least 1,000 battle-related deaths during the war.[23] This is a relatively high threshold of violence for inclusion compared with other sources, as we explain below. The main advantage of COW is that it reports civil wars since 1816, and this long data series allows us to run a simple falsification test using pre-existing trends in conflict. When using COW, we assign the number of years with conflict to the reference dates as follows: wars from 1940-1949 are assigned to 1940, wars from 1950-1959 are assigned to 1950, and so on.[24]

Our second database, covering dates since 1946, is the Uppsala Conflict Data Project, in conjunction with International Peace Research Institute (UCDP/PRIO Armed Conflict

[22]Here we describe our main dependent and independent variables, and a full description of all variables and sources including all controls and baseline characteristics can be found in Appendix Table A-1.

[23]To constitute effective resistance, both sides must have been initially organized for violent conflict, or the weaker side must be able to inflict the opponents at least five percent of the number of fatalities it sustains.

[24]Criteria for inclusion in the COW dataset include a population threshold of 500,000 and having diplomatic recognition (prior to 1920, recognition at or above the rank of *charge d'affaires* with Britain and France and, later, being a member of the League of Nations or the United Nations, or receiving diplomatic missions from two major powers). Costa Rica and Australia are not in the dataset for 1900. While it may be seem reasonable to include them as (peaceful) states in 1900 for our falsification regressions, we avoided making such adjustments to the data, instead followed the choices made by the authors of this and the other codings of civil war.

Dataset Version 4, Gleditsch et al, 2002). We assign number of years in conflict to reference dates as follows: 1946-1949 to 1940, 1950-1959 to 1950, 1960-1969 to 1960, etc. In the case of reference year 1940, we divide the number of years in war by 4 (as the data only start in 1949); for other reference years we divide by 10. This dataset includes conflicts where at least one of the primary parties is the government of a state, and where the use of armed force results in at least 25 battle-related deaths per year. The dataset includes four types of conflicts, and we use the two categories for internal conflict ("internal armed conflict" and "internationalized internal armed conflict").

Our third database is Fearon and Laitin's (2003) coding of civil war. These data cover the period 1945-1999, and the criteria are broadly similar to those of COW[25], except that anticolonial wars are coded as occurring within the empire in question (e.g., Algeria in the 1950s is assigned to France). As with the other datasets, we count the number of years that have any incidence of war, and use our usual rule for assignment to reference dates ($1940 = 1945 - 1949$, $1950 = 1950 - 1959$, etc.).

To examine effects on the intensity of conflict and as a further robustness check, we use information on battle deaths from the Center for the Study of Civil War (CSCW)'s Battle Deaths Dataset (Lacina and Gleditsch, 2005). We use version 3, compatible with the UCDP/PRIO dataset instead of the COW dataset, since the former has a lower threshold of battle-deaths for inclusion and includes more conflicts. This also allows us to more specifically check the robustness of our results in the presence of potential mechanical effects, i.e., to the detection and measurement of civil wars may increase simply because the population is larger and the number of potential deaths is higher. We rely on their "best estimate" of annual battle-related deaths (again we assign deaths to reference years using the rule: $1940 = 1940 - 1949$, $1950 = 1950 - 1959$, etc.)

We have at least partial data for the 65 countries listed in Appendix Table A-1, although we have complete data from 1940 or earlier for only 52 countries (51 when using COW since Austria enters the COW state system in the 1950s). As highlighted previously in footnote 14, we are able to include only five African countries (Algeria, Egypt, Morocco, South Africa, and Tunisia), and this is an important constraint given the prevalence of civil war in Africa. Unfortunately, there is no reliable historical data on causes of death for sub-Saharan Africa

[25]Conflicts are included if they: involved fighting between agents of (or claimants to) a state and organized, nonstate groups who sought control of a government, region, or change in government policies; killed at least 1,000 over its course, with a yearly average of at least 100; at least 100 were killed on both sides (including civilians attacked by rebels).

during the period under investigation.

4.1 Coding Issues

During the post-1940 time period, some countries became independent, others lost their independence, fragmented, or experienced a significant change in borders. For each country, we check when the respective datasets consider the country as entering or leaving the state system, and adjust our measures accordingly. Thus, for example, as Algeria enters the COW system membership in 1962, the measure of conflict for 1960 is the number of years in conflict from 1962-1969 (if any), divided by 8 (instead of 10). We code as missing (not zero) all observations for Algeria in reference years prior to 1960.

As a general rule, for countries that are divided into several states at some point in the sample (e.g., the USSR or Germany), and these embark in *external* wars between them, we do not code them as internal wars of the larger territory. We thus avoid using criteria of our own to define internal conflicts. We do, however, aggregate *internal* wars of member states for such larger countries. Thus, for example, we add USSR internal conflicts while it existed, and aggregate internal conflicts (if any) of the formerly member states and assign them to the USSR as a whole after 1991.[26]

This procedure also minimizes potential mismatches between the level of aggregation of the population figures from Maddison (2006) and civil conflict/political data. Indeed, in the case of Czechoslovakia/Czech Republic, Maddison presents data for Czechoslovakia as a whole, even after split between Slovakia and the Czech Republic. Similarly, population figures are for Vietnam as a whole, and for the USSR while it existed and later the total for ex-USSR.

Maddison's treatment of Germany is more complicated. He takes the 1870 frontiers until 1918, the 1936 frontiers for 1919-1945, and present-day frontiers subsequently. Also, it must be noted that the immediate post-war disease data from the UN are divided into Eastern Germany, Federal Republic of Germany, Berlin, and West Berlin, and numbers for the Federal Republic were used in Acemoglu and Johnson (2007). To make sure our results do not depend on any of these choices, we also dropped Czechoslovakia, Germany, the USSR, and Vietnam and found results similar to those reported below.

[26] These choices make little difference in practice. The countries in our sample potentially affected are just Czechoslovakia, Germany, the USSR and Vietnam. Also, our main specifications end in 1980, prior to many of these splits. Finally, in many cases the dependent variable would be the same aggregating the territories or not. For instance, for the Czech Republic in the 1990s, our dependent conflict variables are always zero with or without aggregating Slovakia.

The construction of our instruments is described fully in Acemoglu and Johnson (2007)[27]. Information on age structure is from the United Nations. We also consider a number of control variables in our robustness exercises, all of which are described in Appendix Table A-1. These include measures of institutions, whether countries were independent in 1940 or not, whether the country was affected by World War II, initial (in 1930) GDP per capita, availability of natural resources (diamonds, oil, and gas), ethnic and religious fragmentation, and the share of Catholic, Muslim, and Protestant populations.

4.2 Descriptive Statistics

Table 1 presents descriptive statistics (sample means and standard deviations) for our baseline sample. We present these summary statistics for the sample as a whole, for groups of countries by income, as well as dividing them between countries experiencing a change in predicted mortality above and below the median. The first ten rows of column 2 show a general trend, evident across all measures, of increasing conflict from the 1940s to the 1980s. Also, columns 3 to 5 show that such an increase is concentrated in middle-income and, especially, poor countries. But more importantly, comparing the change in our conflict measures from 1940 to 1980 in columns 6 and 7, we observe that countries above median change in predicted mortality exhibit larger increases in conflict than those below the median change. For instance, the average years in conflict (per decade) according to the COW measure increased from 0.98 years to 2.09 years for countries with above median change in predicted mortality from 1940 to 1980, while it decreased from 0.44 years to 0.25 years for those with below-median change. This comparison is suggestive for our hypothesis, and we examine below if it survives in our regression exercises and robustness checks.

We measure population in thousands, so an initial population of 1 million is 1,000 in our dataset. We work with log population in order to minimize the effect of outliers, and because average population growth in most countries is better approximated by exponential growth (constant percentage increases) than linear growth (constant absolute increases).

In our base sample, the mean value of log population in 1940 was 9.136 (around 9.3 million), rising to 9.812 in 1980 (i.e., average population doubled to just over 18.2 million).[28]

The change in log population is 0.676.

[27] The main source of the necessary health data on incidence of diseases circa 1940 is the League of Nations (based on national statistics), but other sources were consulted for consistency.

[28] We use the natural logarithm, i.e., the inverse of $\exp(x)$.

4.3 Ordinary Least Squares (OLS) Results

We begin with simple ordinary least squares (OLS) regressions of conflict on population. More specifically, in Table 2 we report regressions of the form,

$$c_{it} = \pi x_{it} + \zeta_i + \mu_t + \mathbf{Z}'_{it}\beta + \varepsilon_{it}, \tag{4}$$

where c_{it} is a measure of conflict for country i and reference year t, and x_{it} is the logarithm of population. ζ_i denotes a full set of country fixed effects while μ_t represents a full set of year dummies; we always include both to remove time-invariant country-specific factors and global trends affecting population and conflict. Z_{it} is a vector of other controls. For all of our regressions, we calculate standard errors that are fully robust against serial correlation at the country level (e.g., as in Wooldridge, 2002, p. 275).[29]

In Table 2, as in subsequent tables, we present two types of estimation: long differences (Panels A and C in Table 2), and panel regressions (Panels B and D). The *long differences* specifications use data only from 1940 (i.e., the 1940s, assigned to 1940) and 1980 (i.e., the 1980s, assigned to 1980). In these specifications, equation (4) is equivalent to a regression of the change in conflict between the two dates on the change in log population between the same two dates, which yields a particularly simple interpretation. Panel regressions use data for intermediate years with one observation per decade (i.e., t = 1940, 1950, 1960, 1970, 1980), and are unbalanced subject only to data availability.

The OLS results in columns 1, 2, and 3 of Table 2 reveal that population is positively correlated with conflict. The estimated coefficient for log population (0.323) in the long-difference regression in column 1 of Panel A, measuring conflict using the COW dataset, implies that the average change in log population in our sample of 0.676 is correlated with about 2.18 more years in conflict in the 1980s relative to the 1940s.[30]

The size of this coefficient is fairly stable across different conflict datasets, as seen in columns 2 and 3, which use the Uppsala and Fearon-Laitin datasets respectively. To address concerns that there may be some mechanical size issue determining what is measured as conflict, column 4 considers log(1+ battle deaths per initial population) as the dependent variable. The

[29] One concern is that these standard errors may be downward biased due to a small number of clusters. Thus, we also implemented the wild bootstrap procedure suggested by Cameron, Gelbach, and Miller (2008). The results, available upon request, are not sensitive to this alternative, which is consistent with our number of clusters being somewhat larger than what typically is considered as small (between 5 and 30 clusters). Cameron, Gelbach, and Miller (2008) also find very similar rejection rates for the cluster robust and wild bootstrap standard errors in their Monte Carlo simulations with 50 clusters.

[30] This is 0.676 multiplied by 0.323, and then multiplied by 10 (as our dependent variable is the fraction of the decade that the country is in conflict).

resulting coefficient for population is also positive and significant at the 90 percent confidence level. Panel B shows similar results from estimating (4) using panel data.

One possible concern with the results in Panels A and B is that they might by driven by age composition effects. In particular, rather than *larger* populations being associated with more civil conflict, it may be that *younger* populations are an important causal factor. For instance, Urdal (2006) finds that exceptionally large youth cohorts, or "youth bulges," correlate with armed conflict, terrorism and rioting. He interprets this as occurring both because of greater opportunities for violence through the abundant supply of youths with low opportunity costs, and stronger motives for violence in societies that cannot respond youth needs. This idea has received considerable attention both in academia and in the general public. As Urdal notes, Huntington (1996) claims that Islam is not any more violent than any other religions, but the demographic factor is key because a high birth rate in the 60s and 70s created a youth bulge in the Muslim world, and people who go out and kill other people are young males.

Panels C and D assess this point with similar regressions as preceding panels, but now with the share of population from 15 to 34 years of age included as an additional independent variable (we lose eight countries due to lack of data). Though the share of young people, which is likely endogenous to population growth, is a 'bad control' (Angrist and Pischke, 2008), this specification is nonetheless a useful to verify whether there is a correlation between population and conflict over and beyond that which would be predicted by the presence of larger young cohorts. Results are consistent with Panel A and B – the coefficients and significance for log population are similar. The point estimate on the share of young population is negative, and significant in the panel regressions of Panel D. At least in this OLS specification, having more young people, once we control for log population, actually reduces conflict.

However, these OLS estimates are not necessarily causal, and the true effect of population on conflict might be larger or smaller than implied by these coefficients. We investigate this issue by applying a plausible instrumental variable.

5 International Epidemiological Transition

Our identification strategy relies on the *International Epidemiological Transition* creating large increases in population. Such increase in populations followed major exogenous (to most countries) innovations in drugs (e.g., penicillin) and associated effective treatments, and chemicals (e.g., DDT). International programs to spread best practices followed through, led by international agencies such as the WHO and UNICEF. This episode provides an instrument for population growth, by using information on the pre-intervention distribution of mortality from various diseases around the world – along with the dates of major global interventions affecting mortality from this set of diseases.

More specifically, we use the *predicted mortality instrument* from Acemoglu and Johnson (2007) which adds each country's initial (in 1940) mortality rate from 15 diseases until there is a global intervention. After the global intervention, the mortality rate from the disease in question declines to the frontier mortality rate[31]. For country i at time t, the instrument is:

$$M_{it}^{I} = \sum_{d \in D} ((1 - I_{dt}) M_{di40} + I_{dt} M_{dFt}), \tag{5}$$

where: M_{di40} denotes mortality in 1940 (measured as number of deaths per 100 individuals per annum) for country i, from disease $d \in D$; I_{dt} is a dummy for intervention for disease d that takes the value of 1 for all dates after the intervention; M_{dFt} is mortality from disease d at the health frontier of the world at time t; and D is the set of diseases listed above.

Since M_{di40} is the pre-intervention mortality rate for disease d, and $I_{dt} = 1$ after a global intervention, the variation in this variable comes from the interaction of baseline cross-country disease prevalence with global intervention dates for those specific diseases. Countries that experienced higher mortality than others for a given disease are expected to observe larger increases in population after the intervention.

The predicted mortality instrument depends on the choice for dating global interventions. An alternative instrument that is independent of the coding of global interventions assumes each country's initial mortality rate decreases at the pace of the global mortality rate for the disease in question. The formula for this *global mortality instrument* is given by,

$$M_{it}^{I} = \sum_{d \in D} \frac{M_{dt}}{M_{d40}} M_{di40}, \tag{6}$$

where M_{dt} is global mortality from disease d in year t, and M_{d40} is global mortality from disease d in 1940, calculated as the unweighted average across countries in the sample of countries in Acemoglu and Johnson (2007).

We use these variables as instruments for population. Specifically, we posit the first-stage relationship for country i at time t,

$$x_{it} = \varphi M_{it}^{I} + \tilde{\zeta}_i + \tilde{\mu}_t + \mathbf{Z}_{it}' \tilde{\beta} + u_{it}, \tag{7}$$

[31] The 15 diseases are (in rough descending order of importance): malaria, pneumonia, and tuberculosis; influenza, cholera, typhoid, smallpox, shigella dysentery, whooping cough, measles (rubeola), dyphteria, scarlet fever, yellow fever, plague, typhus.

where: x_{it} is the logarithm of population; M_{it}^{I} the predicted (or global) mortality instrument; $\tilde{\zeta}_i$ is a full set of country fixed effects; $\tilde{\mu}_t$ are year fixed effects; and $\mathbf{Z}_{it}$ represents a vector of other controls.

Acemoglu and Johnson (2007) show that the changes in predicted mortality led to major improvements in life expectancy and other measures of health. In countries such as India, Pakistan, Indonesia, Ecuador and El Salvador, where predicted mortality declined by a large amount, there were large gains in life expectancy. Instead, life expectancy remained comparatively unchanged in parts of western Europe, Uruguay, Argentina, Korea, and Australia, where predicted mortality did not decrease as much. The same negative relationship holds without the richest countries, so it is not driven by the comparison of initially rich countries to initially low- and middle-income countries.

6 Main Results

6.1 First Stages

Table 3 shows the first-stage relationship, i.e., estimating equation (7). This table shows the strong negative relationship between log population and predicted mortality is robust across alternative samples. Panel A reports long-difference specifications, and panel B reports panel regressions.

Column 1 includes all countries in our sample, and shows an estimate of φ equal to -0.782, which is significant at less than 1 percent. This estimate implies that an improvement in predicted mortality of 0.469 per 100 (or 469 per $100,000$, which is the mean improvement between 1940 and 1980 in our base sample) leads to an increase of roughly 0.37 in log population – thus close to a 37% increase in total population. The mean population in our sample in 1940 was about 34.7 million, so this is an increase of roughly 12.8 million, whereas the actual mean increase in population between 1940 and 1980 was about 23.5 million. This implies that changes in predicted mortality account for approximately one-half of the increase in population between 1940 and 1980.

Column 2 repeats the same regression excluding Eastern Europe, and Column 3 looks only at initially low- and middle-income countries. The estimate of φ is similar, and still significant at less than 1 percent. Column 4 presents results using the global mortality instrument. The results are also strong and significant, reassuring us that they do not depend on the coding of global intervention dates. Finally, column 5 excludes the countries most affected by World

War II, again with almost identical results.

Panel B repeats the same regressions as in Panel A, now using a panel with decadal observations. The results are still highly significant but the coefficients are smaller, which is reasonable since these regressions exploit shorter-run responses to changes in predicted mortality.

6.2 Robustness to Differential Trends

The main potential threat to our exclusion restriction would be that the 1940 mortality rates are somehow correlated with future changes in conflict. We therefore need to examine the robustness of our IV results to the inclusion of differential trends that are parametrized as functions of various baseline characteristics. Whether this explains the first-stage relationship is investigated with regressions of the form,

$$x_{it} = \varphi M_{it}^{I} + \tilde{\zeta}_i + \tilde{\mu}_t + \sum_{t=1940}^{1980} \kappa_i' \bar{\omega}_t + u_{it}, \tag{8}$$

where $\bar{\omega}_t = 1$ in year t and zero otherwise, and κ_i are "time-invariant" characteristics of country i. These characteristics include: a measure of the average quality of institutions (average of the constraints on the executive from the Polity IV data set over 1950-70); a dummy for the country being independent in 1940; initial (in 1930) GDP per capita, population, and share of young people; and measures of the availability of natural resources and ethnic polarization/fragmentation, which are often emphasized on the empirical literature on civil war. These regressions are reported in Table 4.

Since equation (8) includes a full set of time interactions with κ_i, we are controlling for differential trends related to these characteristics. In long-difference regressions of panel A, this specification is equivalent to including an interaction between the 1980 dummy and the various baseline characteristics.

The results in both panels show that controlling for these characteristics has little effect on our results. The coefficient on predicted mortality remains negative and significant across all columns. Overall, the instrument is strong and its correlation with population is unlikely to be driven by differential trends due to a third factor.

6.3 Reduced Forms and Falsification

There is no evidence of a negative relationship between pre-existing trends in life expectancy and subsequent changes in predicted mortality (if anything, the relationship is slightly positive).[32] There is no clear correlation between prior changes in population and changes in predicted mortality. This stands in sharp contrast with the correlation between predicted mortality and population observed after 1940.

Table 5 (and Figure 4) reports the results of reduced form regressions and falsification tests. We run the following type of regression,

$$\Delta y_{it_1,t_0} = \alpha + \varphi \Delta M^I_{i1980,1940} + \varepsilon_{it}. \tag{9}$$

where α is a constant and $\Delta y_{it_1,t_0} \equiv y_{it_1} - y_{it_0}$ is the change in our dependent variable for country i between reference dates t_0 and t_1. Similarly, $\Delta M^I_{i1980,1940} \equiv y_{i,1980} - y_{i,1940}$ is the change in the predicted mortality instrument between 1940 and 1980.

In columns 1 and 2, the dependent variable is the change in the fraction of each decade in conflict from 1940 to 1980. Notice that this specification is equivalent to a long-difference regression (using only data for 1940s and 1980s) of conflict on predicted mortality with a full set of country fixed effects. It is therefore the reduced-form regression for our simplest long-difference specification[33]. These columns, for the base sample and for low- and middle-income countries, show that countries with a larger decline in predicted mortality experienced a larger increase in years in conflict. Given the negative relationship between predicted mortality and population shown in the previous section, this translates into a positive effect of population on conflict in our 2SLS estimates below.

A useful falsification exercise is to look at changes in predicted mortality, and see whether they correlate with changes in conflict or population during the pre-period. That is, we consider specifications of equation (9) where $t_1 = 1940$ and $t_0 = 1900$. In Columns 3 and 4 of Table 5, we find no relationship between the change in conflict from 1900 to 1940 and change in predicted mortality from 1940 to 1980, for the base sample and for low- and middle-income countries. Similar specifications for our first stage (changes in log population from 1900 to 1940 and in predicted mortality from 1940 to 1980) are shown in columns 5 and 6, again with no sign of such a relationship. Predicted mortality explains changes in population after 1940, but not before 1940. The coefficient estimates are very small relative to our reduced forms, not just insignificant.

[32] These issues are examined in greater detail in Acemoglu and Johnson (2007). Their results suggest a robust and significant relationship between predicted mortality and health that is unlikely to be driven by preexisting trends.

[33] In the Appendix, we show the panel versions of these reduced forms, as well as their robustness to inclusion of differential trends in Tables A-2, A-3 and A-7.

These results offer further confirmation there were no preexisting trends related to changes in predicted mortality either in population or in our key conflict outcome variables. This gives us greater confidence in using predicted mortality as an instrument to investigate the effect of population on conflict.

6.4 2SLS Results

Table 6 presents our main results, which are the 2SLS estimates of the effect of population on conflict. More specifically, our second stage regression is given by equation (4), where population is instrumented by predicted mortality –equation (7). As before, we report long-difference regressions for 1940 and 1980 in panel A and panel regressions for $1940 - 1980$ in panel B. This table shows that the effect of population on conflict is positive, and very significant in most specifications.[34]

In column 1, the dependent variable is the share of years in internal conflict per decade, as measured by the COW dataset. The size of the effect (π) is estimated to be 0.617, which implies that the average change (0.676) in log population from 1940 to 1980 leads approximately to 4.17 more years in conflict during the 1980s relative to the 1940s.

This can be compared to the OLS coefficient in Table 2 (0.323), which implied an effect of around 2.18 more years in conflict in the 1980s compared to the 1940s. We find similar results in the case of the panel regressions for 1940-80 presented in panel B ($\pi = 0.57$, significant at the 99% level).

For a country like El Salvador, experiencing an increase in population from 1.6 to 4.6 million in this period (a change in log population of 0.46), the OLS estimate predicts roughly 1.5 ($0.323 \times 0.46 \times 10$) more years in conflict per decade while the IV estimate of 0.617 implies an effect of roughly 2.8 more years in conflict ($0.617 \times 0.46 \times 10$).

Columns 2 through 5 investigate the robustness of this result. The dependent variables in columns 2 and 3 are the years in internal conflict as a fraction of total years in the reference date as measured by the UCDP/PRIO and Fearon and Laitin datasets, respectively. All the estimated coefficients are positive, and typically significant at less than 1 or 5%, with the exception of the UCDP/PRIO regressions in Panel B.

Since conventional measures of civil war rely on meeting a battle death threshold, an increase in total population may mechanically increase the number of "detected" civil wars.

[34] The exclusion restriction for our IV strategy -$Cov(M^I_{it}, \varepsilon_{it}) = 0$, where ε_{it} is the error term in the second-stage equation, requires that the unique channel for casual effects of predicted mortality on conflict is changes in population. This does not seem unreasonable.

We use battle deaths data to examine whether this may be driving our results. Column 4 considers the (log of) battle deaths for each reference date, per person, to calculate c_{it}. The coefficient on population is also positive and significant. Finally, columns 5 through 8 repeat the regressions from columns 1 through 5 but use global mortality as the instrument for population. The results are very similar. This evidence suggests that our results do not depend on the dating of global health interventions.

7 Robustness Checks

7.1 Controlling for Differential Trends

An important potential threat to our strategy is that our estimated causal effects of population on conflict could be actually capturing differential trends between countries which happen to have different levels of baseline mortality rates. We therefore need to examine the robustness of our results to the inclusion of differential trends, parametrized as functions of various observable baseline characteristics. In choosing these characteristics, we draw on the extensive literature on civil wars.

In Table 7, in line with the corresponding first stages in equation (8) and Table 4, our second stage equations take the following form:

$$c_{it} = \pi x_{it} + \zeta_i + \mu_t + \sum_{t=1940}^{1980} \kappa_i' \bar{\omega}_t + \varepsilon_{it}. \tag{10}$$

In column 1, we examine whether the results could be driven by differential trends between countries with "good" and "bad" institutions. While there are many dimensions of institutions, we choose to measure the quality of institutions by average constraints on the executive over 1950-1970. This is a particularly relevant dimension of institutions, since, as noted in Section 3, the commitment problem is a persuasive explanation for civil war. In column 2, κ_i is simply a dummy variable equal to 1 if country i was independent in 1940. Columns 3 to 5 control for differential trends as a function of initial (1930) log GDP per capita, initial log population, and initial share of young (population aged 15 to 34), respectively.

In columns 6 through 9, the country characteristics κ_i are variables emphasized by other researchers as correlates of civil war. A large literature links conflict to natural-resource abundance, in particular oil, gas, and diamonds. A commonly used measure is oil exports divided by GDP or the share of the natural resource sector in GDP (Sachs and Warner, 1995). As Ross (2006) notes, this measure may be a poor proxy of rents in the economy or potential revenues for the government since it does not include oil that is produced but consumed domestically, and it does not account for extraction costs which may vary across countries. Also, even at similar levels of production, the numerator tends to be larger in poor countries because poor countries consume less of their own oil. Normalizing by GDP similarly inflates the numbers for poor countries. Motivated by this reasoning, in columns 6 and 7, κ_i is, respectively: diamond production per capita (from Humphreys, 2005), and oil and gas rents per capita (from Ross, 2006).[35]

A number of theories also suggest that ethnic (or religious) diversity and polarization may be a cause civil war, or at least that they may facilitate surmount the big collective action problems within groups in conflict. Nevertheless, cross-national studies find few differences between the determinants of civil war in general versus "ethnic" civil wars in particular (see Fearon (2006) for a review). This may be surprising, yet it could be driven by the fact that ethnic fragmentation is measured with considerable error. As Blattman and Miguel (2010) point out, the existing proxies may also be theoretically inappropriate and these indices of ethnic fractionalization have been questioned as a meaningful proxy for ethnic tensions (e.g., Posner 2004a, 2004b). Esteban and Ray (1994, 1999) argue that more than fractionalization, a bimodal distribution of preferences or resources—"polarization"—is linked to greater conflict risk. Montalvo and Reynal-Querol (2005) construct measures for polarization and fragmentation and find support for this theory. In columns 8 and 9 we use their measures of ethnic polarization and fractionalization.

Notice that the coefficient remains significant at conventional confidence levels in every regression. Similarly, the panel regressions suggest a significant positive effect. Moreover, the coefficient is quite stable across specifications, ranging from around 0.6 to 0.75 in most long-difference specifications. The sole exception is column 3, which includes a differential trend by initial GDP per capita; here the estimated coefficient increases to 1.1. This result suggests our estimated impact of population on conflict is unlikely to be explained by differential trends by levels of income. Overall, in fact, Table 7 suggests that it is unlikely that the impact of population on conflict from our 2SLS is actually driven by differential trends[36].

[35]In addition, we found similar results controlling for oil production per capita (also from Humphreys, 2005).

[36]Moreover, while Table 7 uses the best available measures of resource abundance and the more theoretically-motivated measures of ethnic diversity, the results do not depend on the exact variable used to measure natural resource abundance or social diversity. This is verified in Appendix Tables A-5 and A-6, which present the first and second stages, respectively, for specifications similar to those in Table 7 but where alternative measures are used, including: the share of the natural resources in GDP, total (instead of per capita) oil and diamond production, religious polarization and fragmentation, and share of Catholic, Muslim, and Protestant population.

7.2 Alternative Samples, Instrument, and World War II

Table 8 presents additional robustness checks on our main results. To facilitate comparisons, column 1 reproduces our base sample long-difference and panel regression estimates from Table 6. In column 2, we exclude East European countries, which may have exhibited special behavior in the context of the Cold War. The estimated value of π remains positive, of similar size and statistically significant. Column 3 drops initially rich countries to verify that these results are not driven by the comparison between rich and poor nations, and Column 4 uses the global mortality instrument.

Columns 5 through 7 check whether results are driven by events around World War II. Column 5 excludes the countries demographically most affected by the War, namely Austria, China, Finland, Germany, Italy, Russian Federation (Urlanis, 2003). Column 6 assigns instead the level of conflict of the 1950s to the 1940s. Column 7 simply ignores the war years, and assigns the number of years in conflict from 1946-49 (as a fraction of the 4 years in these interval) to our dependent variable in 1940.

Finally, column 8 controls for the share of young (15-34) population, finding similar effects[37]. Overall, the coefficient is very stable and retains statistical significance at conventional levels. These robustness checks thus lend more credibility to our baseline estimates.

7.3 Timing

Table 9 examines how the response of conflict to population growth changed over time. In particular, columns 1 to 5 look at different time horizons by estimating long-difference regressions for our baseline measure of conflict on population (instrumented with predicted mortality), where the initial time period is $t = 1940$ and the final date is 1960, 1970, 1980, 1990, and 2000. Consistent with the idea that health improvements and population increase have a lagged effect on social conflict, as the resulting scarcity finally derives in violence, results are weaker if we only look only at 1940-60 or 1940-70, and the effect peaks in 1980. There is also a significant impact when comparing the 1940s to the 1990s and 200s, though the size of the effect is about a third and 50 percent smaller, respectively, of that for the 1980s. One conjecture is that the nature of a number of conflicts changed with the fall of the Soviet Union and the wave of democratizations of the 1990s. These findings are again not sensitive to the coding of global

[37]We reiterate that this regression must be interpreted with caution since the share of young population is a 'bad control' potentially influenced by the increase in overall population. However, in Appendix Table A-4 we show that predicted mortality does not influence the growth in the share of young population from the 1940s to the 1980s or 1990s.

health interventions, as Panel B reveals.

8 Conclusions

The large and largely unprecedented population increases that followed the *international epidemiological transition* of the 1940s contributed to an increase in violent social conflict. At least in this important historical episode, increasing population without a corresponding increase in resources and technology, raised the likelihood of civil war – presumably because there was an more intense competition for scarce resources.

The international epidemiological transition produced significant convergence in health conditions around the world, but no comparable convergence has been observed in income per capita. At least in part, this lack of convergence for prosperity can be attributed to the negative consequences of social conflict.

The extent to which this historical experience applies to the modern situation remains to be seen. We should expect higher population in some countries that are currently relatively low income. In part these increases are driven by health improvements that have already taken place. Further interventions from the outside are likely to improve life expectancy and further increase population.

The world tendency towards violence in some average sense may have declined, and the potential for growth in low income countries may now be higher than in the past, for example because of changes in technology or better policy. Or perhaps outside interventions will increase productivity and shift people away from having to compete for scarce local resources. But experience from the 1940s-1980s period should at least serve as a cautionary tale – encouraging experts and policy makers to ensure that economic opportunities increase in line with the number of people seeking employment and income.

References

Abadie, A. & J. Gardeazabal (2003). The Economic Costs of Conflict: A Case Study of the Basque Country. American Economic Review, 93(1), 113–32.

Acemoglu, D., Egorov, G., & Sonin, K. (2008). Dynamics and stability of constitutions, coalitions, and clubs. (NBER Working Paper #14239)

Acemoglu, D., & Johnson, S. (2007). Disease and development: The effect of life expectancy on economic growth. Journal of Political Economy, 115(6), 925-985.

Acemoglu, D., & Robinson, J. A. (2001). A theory of political transitions. American Economic Review, 91(4), 938-963.

Acemoglu, D., & Robinson, J. A. (2006). Economic origins of dictatorship and democracy. New York, NY: Cambridge University Press.

Alesina, A., & Ferrara, E. L. (2000). Participation in heterogeneous communities. The Quarterly Journal of Economics, 115(3), 847-904.

Alesina, A., Baqir, R., & Easterly, W. (1999). Public goods and ethnic divi- sions. The Quarterly Journal of Economics, 114(4), 1243–1284.

Angrist, J. D. & J.S. Pischke (2008). Mostly Harmless Econometrics: An Empiricist's Companion. Princeton: Princeton University Press.

Bates, R. H. (1986). Modernization, ethnic competition, and the rationality of politics in contemporary Africa. In M. E. Doro & N. M. Schultz (Eds.), Governing Black Africa. New York and London: Africana Publishing Company.

Besley, T. J., & Persson, T. (2008). The incidence of civil war: Theory and evidence (NBER Working Papers No. 14585). National Bureau of Economic Research, Inc.

Blattman, C., & Miguel, E. (2010). Civil war. Journal of Economic Literature, 48(1), 3-57.

Brückner, M. (2010). Population size and civil conflict risk: Is there a causal link? Economic Journal, 120(544), 535-550.

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. The Review of Economics and Statistics, 90(3), 414-427.

Caselli, F., & II, W. J. C. (2006). On the theory of ethnic conflict (Tech. Rep.).

Collier, P., & Hoeffler, A. (1998). On economic causes of civil war. Oxford Economic Papers, 50(4), 563-73.

Collier, P., & Hoeffler, A. (2004). Greed and grievance in civil war. Oxford Economic Papers, 56(4), 563-595.

Dal Bó, P., & Dal Bó, E. (2004). Workers, warriors and criminals: Social conflict in general equilibrium (Econometric Society 2004 Latin American Meetings No. 341). Econometric Society.

Diehl, P. F., & Gleditsch, N. P. (2001). Environmental conflict. Boulder, CO: Westview Press.

Dube, O., & Vargas, J. F. (2008). Commodity price shocks and civil conflict: Evidence from Colombia. (Harvard University Center for International Development Graduate Student and Postdoctoral Fellow Working Paper 14)

Easterly, W., & Levine, R. (1997). Africa's growth tragedy: Policies and ethnic divisions. Quarterly Journal of Economics, 112(4), 1203-1250.

Esteban, J., & Ray, D. (1999). Conflict and distribution. Journal of Economic Theory, 87(2), 379-415.

Esteban, J., & Ray, D. (2008). On the salience of ethnic conflict. American Economic Review, 98(5), 2185-2202.

Fearon, J. D. (1998). Commitment problems and the spread of ethnic conflict. In D. Lake & D. Rothchild (Eds.), The international spread of ethnic conflict: Fear, diffusion, and escalation. Princeton: Princeton University Press.

Fearon, J. D. (2004). Why do some civil wars last so much longer than others? Journal of Peace Research, 41(3), 275-301.

Fearon, J. D. (2006). Ethnic mobilization and ethnic violence. In B. R. Weingast & D. A. Wittman (Eds.), The oxford handbook of political economy (pp. 852–68). Oxford and New York: Oxford University Press.

Fearon, J. D., & Laitin, D. D. (1996). Explaining interethnic cooperation. The American Political Science Review, 90(4), pp. 715-735.

Fearon, J. D., & Laitin, D. D. (2003). Ethnicity, insurgency, and civil war. The American Political Science Review, 97(1), pp. 75-90.

Fergusson, L., Ibáñez, A.M. & Riaño, J.F. (2015). Conflict, Educational Attainment and Structural Transformation: La Violencia in Colombia, Universidad de los Andes, Documento Cede 2015-35.

Garfinkel, M. R., & Skaperdas, S. (2007). Economics of conflict: An overview. In K. Hartley & T. Sandler (Eds.), Handbook of defense economics (Vol. 2, p. 649-709). Elsevier.

Gates, B. & Gates, M. (2015), "2015 Gates Annual Letter: Our Big Bet for the Future", Avaliable at `http://www.gatesnotes.com/2015-annual-letter`, last accessed May 13 2015.

Gellner, E. (1983). Nations and nationalism. Ithaca and London: Cornell University Press.

Gleditsch, N. P. (1998). Armed conflict and the environment: A critique of the literature. Journal of Peace Research, 35(3), 381-400.

Gleditsch, N. P., Wallensteen, P., Eriksson, M., Sollenberg, M., & Strand, H. (2002). Armed conflict 1946-2001: A new dataset. Journal of Peace Research, 39(5), 615-637.

Hauge, W., & Ellingsen, T. (1998). Beyond environmental scarcity: Causal pathways to conflict. Journal of Peace Research, 35(3), 299-317.

Homer-Dixon, T. F. (1991). On the threshold: Environmental changes as causes of acute conflict. International Security(2), 76 - 116.

Homer-Dixon, T. F. (1994). Environmental scarcities and violent conflict: Evidence from cases. International Security, 19 (1), 5 - 40.

Homer-Dixon, T. F. (1999). Environment, scarcity, and violence. Princeton, NJ: Princeton University Press.

Horowitz, D. L. (1985). Ethnic groups in conflict. Berkeley and London: University of California.

Humphreys, M. (2005). Natural resources, conflict, and conflict resolution. Journal of Conflict Resolution, 49(4), 508-537.

Huntington, S. P. (1996) The Clash of Civilizations and the Remaking of World Order. New York: Simon & Schuster.

Kaplan, R. (1994). The coming anarchy. The Atlantic (February).

Lacina, B., & Gleditsch, N. P. (2005). Monitoring trends in global combat: A new dataset of battle deaths. European Journal of Population, 21(2-3), 145-166.

La Porta, R., Silanes, F. Lopez-de, Shleifer, A., & Vishny, R. (1999). The quality of government. Journal of Law, Economics and Organization, 15(1), 222-79.

MacKay, A. (1981). Climate and popular unrest in late medieval Castile. In M. I. T.M. Wigley & G. Farmer (Eds.), Climate and history: Studies in past climates and their impact on man (p. 356-376). Cambridge: Cambridge University Press.

Maddison, A. (2006). The world economy. Paris, France: Development Centre of the Organisation for Economic Co-operation and Development.

McEvedy, C. & Jones, R. (1978), Atlas of World Population History, New York: Penguin.

Meredith, M. (2005). The fate of africa. New York, NY: Public Affairs.

Miguel, E., & Gugerty, M. K. (2005). Ethnic diversity, social sanctions, and public goods in Kenya. Journal of Public Economics, 89(11-12), 2325-2368.

Miguel, E., Satyanath, S., & Sergenti, E. (2004). Economic shocks and civil conflict: An instrumental variables approach. Journal of Political Economy, 112(4), 725–753.

Montalvo, J. G., & Reynal-Querol, M. (2005). Ethnic polarization, potential conflict, and civil wars. American Economic Review, 95(3), 796-816.

Pinker, S. (2011), The Better Angels of Our Nature: Why Violence Has Declined, New York: Viking.

Posner, D. N. (2004a). Measuring ethnic fractionalization in Africa. American Journal of Political Science, 48(4), pp. 849-863.

Posner, D. N. (2004b). The political salience of cultural difference: Why Chewas and Tumbukas are allies in Zambia and adversaries in Malawi. The American Political Science Review, 98(4), pp. 529-545.

Powell, R. (2006). War as a commitment problem. International Organization , 60 (01), 169-203.

Richards, P. (1996). Fighting for the rain forest: War, youth and resources in sierra leone. The International African Institute.

Ross, M. (2006). A closer look at oil, diamonds, and civil war. Annual Review of Political Science, 9, 265-300.

Ross, M. L. (2008). Oil, Islam, and women. American Political Science Review, 102(01), 107-123.

Sachs, J. D., & Warner, A. M. (1995). Natural resource abundance and economic growth (Working Paper No. 5398). National Bureau of Economic Research.

Sambanis, N. (2002). A review of recent advances and future directions in the quantitative literature on civil war. Defence and Peace Economics, 13(3), 215-243.

Sarkees, M. R., & Schafer, P. (2000). The correlates of war data on war: an update to 1997. Conflict Management and Peace Science, 18(1), 123-144.

Tir, J., & Diehl, P. F. (1998). Demographic pressure and interstate conflict: Linking population growth and density to militarized disputes and wars, 1930-89. Journal of Peace Research, 35(3), 319-339.

Turner, M. E. (1984). Enclosures in britain, 1750-1830. London: Macmillan.

United Nations (2013), "World Population Prospects: The 2012 Revision. Volume 1, Comprehensive Tables," New York: United Nations. Available at `http://esa.un.org/wpp/Documentation/pdf/WPP2012_Volume-I_Comprehensive-Tables.pdf`, last accessed May 13 2015.

United Nations (2011), "World Population Prospects: The 2010 Revision. Volume 1, Comprehensive Tables," New York: United Nations. Available at `http://esa.un.org/wpp/documentation/pdf/WPP2010_Volume-I_Comprehensive-Tables.pdf`, last accessed May 13 2015.

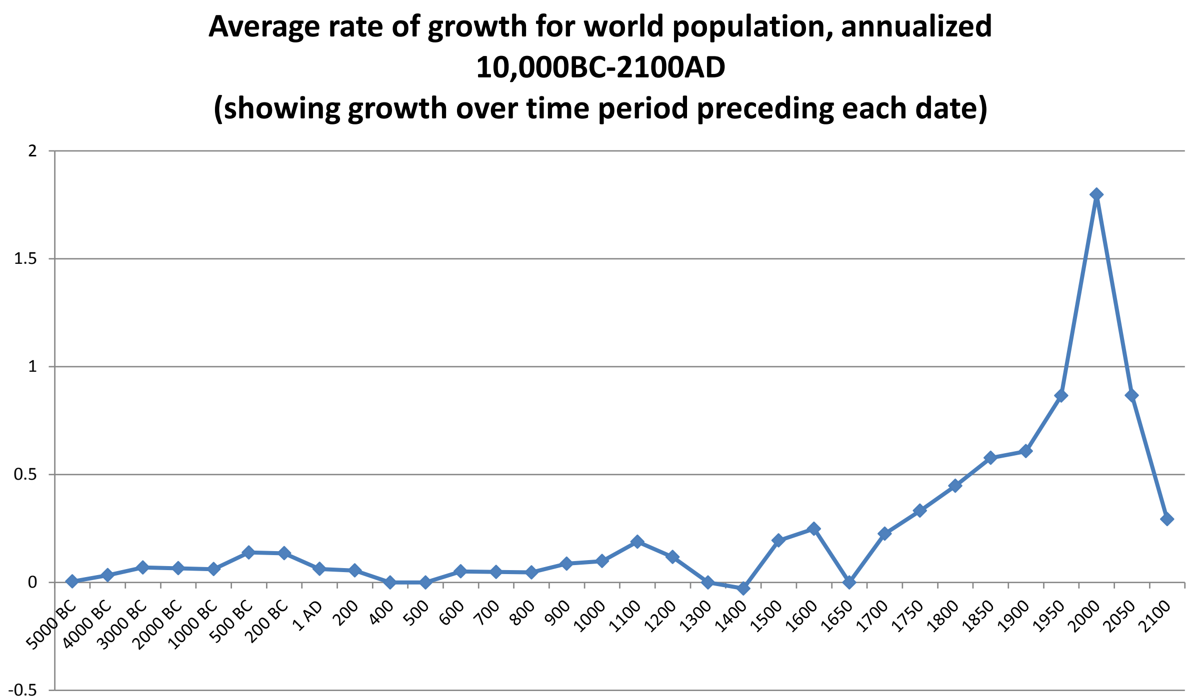
Urdal, H. (2006), "A Clash of Generations? Youth Bulges and Political Violence", International Studies Quarterly, 50(3): 607–629.

Urlanis, B. (2003). Wars and population. Honolulu, Hawaii: University Press of the Pacific. (reprint of 1971 edition)

US Census Bureau (2015), International Data Base IDB, Available at `http://www.census.gov/population/international/data/idb/informationGateway.php`, last accessed May 13 2015.

Wooldridge, J. (2002), Econometric analysis of cross section and panel data. MIT Press.

Figure 1



World population: 4m in 10,000BC; 170m in 1AD; 425m in 1500; 900m in 1800; 1.6bn in 1900; 2.5bn in 1950; 7bn in 2012; 9bn in 2042; and 10.8bn in 2100. World population growth rate peaked in late 1980s.

Figure 2. Population, Life Expectancy and GDP

Initially Rich, Middle-Income and Poor Countries, Base Sample

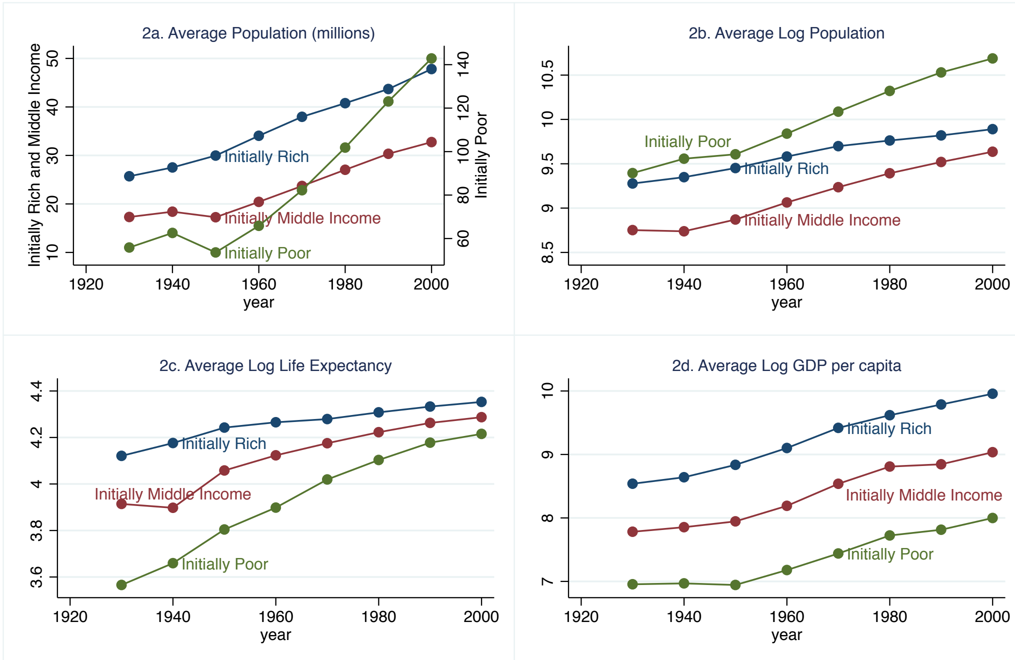


Figure 3. Conflict incidence

Initially Rich, Middle-Income and Poor Countries, Base Sample

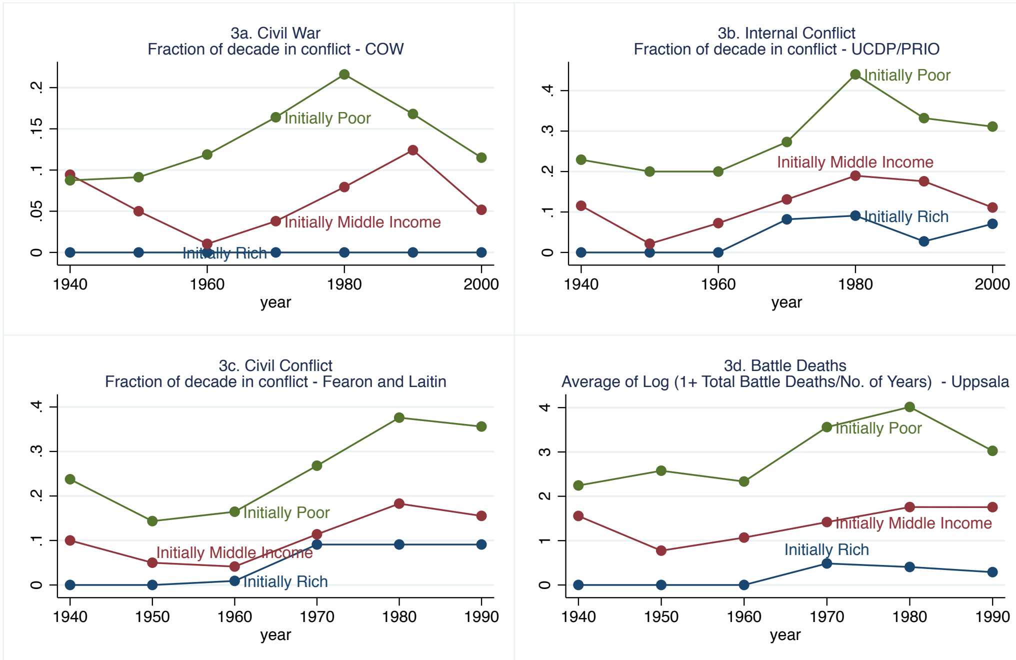


Figure 4. Reduced form and falsification

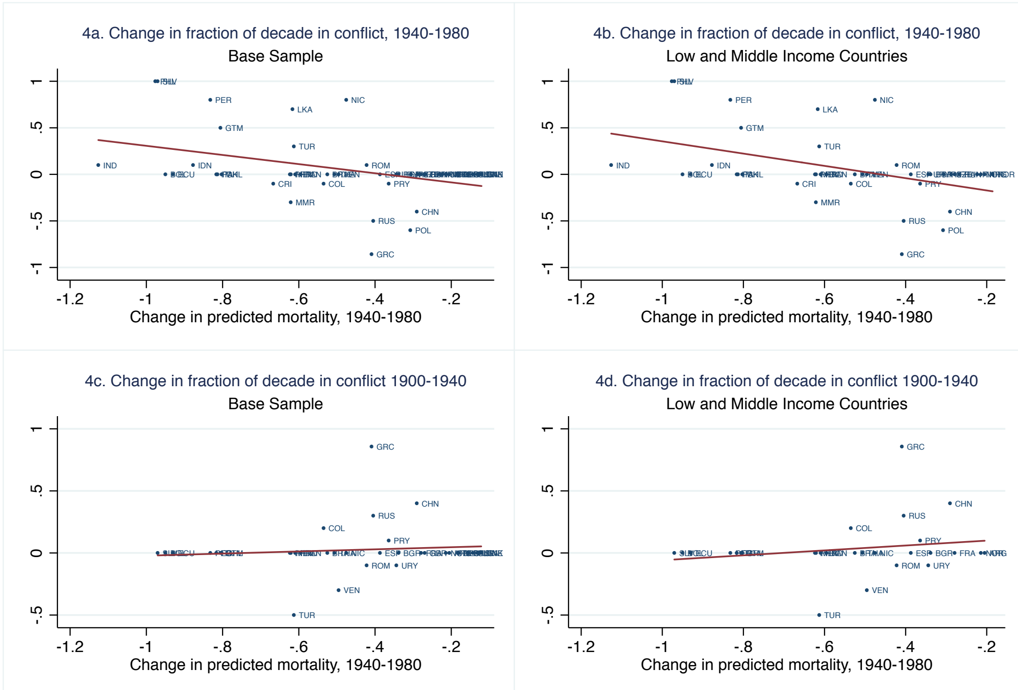


Table 1: Descriptive statistics

			By initial income			By change in predicted mortality	
	Year	Base sample	Rich	Middle	Poor	Above median	Below median
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Fraction of decade in conflict, COW	1940	0.072	0	0.094	0.088	0.098	0.044
		(0.212)	(0)	(0.222)	(0.263)	(0.261)	(0.142)
Fraction of decade in conflict, COW	1980	0.118	0	0.079	0.216	0.209	0.025
		(0.260)	(0)	(0.188)	(0.347)	(0.328)	(0.102)
Fraction of decade in conflict, Uppsala	1940	0.126	0	0.115	0.229	0.201	0.048
		(0.306)	(0)	(0.276)	(0.417)	(0.369)	(0.200)
Fraction of decade in conflict, Uppsala	1980	0.269	0.091	0.190	0.440	0.400	0.134
		(0.408)	(0.302)	(0.341)	(0.468)	(0.447)	(0.319)
Fraction of decade in conflict, Fearon-Laitin	1940	0.121	0	0.100	0.238	0.163	0.077
		(0.302)	(0)	(0.255)	(0.427)	(0.347)	(0.247)
Fraction of decade in conflict, Fearon-Laitin	1980	0.242	0.091	0.183	0.376	0.367	0.112
		(0.413)	(0.302)	(0.369)	(0.475)	(0.467)	(0.306)
Log 1 + Battle Deaths/years, Uppsala	1940	1.441	0	1.556	2.244	2.105	0.751
		(3.232)	(0)	(3.159)	(4.165)	(3.545)	(2.772)
Log 1 + Battle Deaths/years, Uppsala	1980	2.396	0.405	1.757	4.014	3.696	1.056
		(3.321)	(1.344)	(2.907)	(3.710)	(3.672)	(2.280)
Log 1 + Battle Deaths/population, Uppsala	1940	0.157	0	0.251	0.112	0.208	0.105
		(0.527)	(0)	(0.707)	(0.305)	(0.644)	(0.373)
Log 1 + Battle Deaths/population, Uppsala	1980	0.282	0.002	0.159	0.638	0.541	0.005
		(0.779)	(0.005)	(0.528)	(1.153)	(1.024)	(0.017)
Log of population	1940	9.136	9.349	8.738	9.557	9.010	9.261
		(1.455)	(1.344)	(1.192)	(1.756)	(1.530)	(1.393)
Log of population	1980	9.812	9.762	9.393	10.321	9.856	9.768
		(1.384)	(1.293)	(1.238)	(1.461)	(1.484)	(1.294)
Baseline predicted mortality	1940	0.469	0.171	0.487	0.626	0.690	0.241
		(0.271)	(0.050)	(0.224)	(0.272)	(0.195)	(0.080)
Global predicted mortality	1940	0.456	0.171	0.482	0.593	0.666	0.238
		(0.258)	(0.050)	(0.222)	(0.252)	(0.184)	(0.079)

Notes: The table reports the mean values of variables in the samples described in the column heading, with standard deviations in parentheses. Initially rich countries had log GDP per capita over 8.4 in 1940, middle-income countries had log GDP per capita between 7.37 and 8.4, and low-income countries had log GDP per capita below 7.37 in 1940. Predicted mortality is measured per 100 per year. Columns 6 and 7 report descriptive statistics for subsamples in which change in predicted mortality between 1940 and 1980s was above or below the median value in the base sample (-0.405). Initially rich countries have no civil wars recorded in the COW dataset in the 1940s and 1980s, and no conflict incidence according to the Fearon and Laitin and Uppsala sources in the 1940s. See the text and Appendix Table 1 for details and definitions.

Table 2: Population and Conflict: OLS Estimates

Dependent variable...	Fraction of decade in conflict			Log(1+Battle Deaths/ Pop. 1940)
	COW	Uppsala	Fearon & Laitin	
	(1)	(2)	(3)	(4)
Panel A: long differences, just 1940s and 1980s				
log of population	0.323***	0.271**	0.236*	0.722*
	(0.116)	(0.135)	(0.139)	(0.401)
Observations	102	104	104	104
R-squared	0.177	0.145	0.098	0.102
Number of clusters	50	51	51	51
Panel B: panel regressions, 1940s-1980s				
log of population	0.268***	0.311**	0.251*	0.738*
	(0.095)	(0.132)	(0.132)	(0.375)
Observations	307	308	308	273
R-squared	0.086	0.146	0.113	0.106
Number of clusters	63	63	63	54
Panel C: long differences controlling for age structure, just 1940s and 1980s				
log of population	0.391***	0.316**	0.344**	1.043**
	(0.140)	(0.149)	(0.162)	(0.469)
share of population 15-34	-0.995	-0.529	-3.504	-4.852
	(1.271)	(1.544)	(2.805)	(4.678)
Observations	86	88	88	88
R-squared	0.226	0.222	0.178	0.193
Number of countrynum	43	44	44	44
Number of clusters	43	44	44	44
Panel D: panel regressions controlling for age structure, 1940s-1980s				
log of population	0.331**	0.313**	0.265*	0.968**
	(0.124)	(0.153)	(0.148)	(0.442)
share of population 15-34	-1.068**	-1.050*	-2.095	-4.102**
	(0.490)	(0.599)	(1.432)	(1.858)
Observations	227	228	228	228
R-squared	0.157	0.161	0.131	0.171
Number of clusters	46	46	46	46

Notes: OLS regressions with a full set of year and country fixed effects (equation (4) in the text). Robust standard errors (clustered by country) are reported in parentheses. Panels A and C are long-difference specifications with two observations per country, one for the initial date and one for the final date. Panels B and D are unbalanced panels with one observation per decade. For clustered standard errors, Bangladesh, India and Pakistan are considered a single cluster. See the text and Appendix Table A-1 for definitions and details.

Table 3: Predicted Mortality and Population:
First Stage Estimates and Basic Robustness

	Dependent variable is log of population				
	Base Sample	Excluding Eastern Europe	Low and Middle Income Countries Only	Global Mortality Instrument	Excluding Most Affected By WWII
	(1)	(2)	(3)	(4)	(5)
Panel A: long differences, just 1940s and 1980s					
baseline predicted mortality	-0.782***	-0.700***	-0.764***	-0.818***	-0.811***
	(0.141)	(0.141)	(0.191)	(0.145)	(0.140)
Observations	102	92	80	102	94
R-squared	0.823	0.847	0.828	0.823	0.842
Number of clusters	50	45	39	50	46
Panel B: panel regressions, 1940s-1980s					
baseline predicted mortality	-0.464***	-0.402***	-0.471***	-0.681***	-0.476***
	(0.094)	(0.093)	(0.131)	(0.128)	(0.095)
Observations	307	278	252	307	279
R-squared	0.792	0.819	0.814	0.814	0.815
Number of clusters	63	57	52	63	57

Notes: OLS regressions with a full set of year and country fixed effects (equation (7) in the text). Robust standard errors (clustered by country) are reported in parentheses. Panel A presents long-difference specifications with two observations per country, one for the initial date and one for the final date. Panel B presents unbalanced panels with one observation per decade. For clustered standard errors, Bangladesh, India and Pakistan are considered a single cluster. See the text and Appendix Table A-1 for definitions and details.

Table 4: Predicted Mortality and Population: First Stage Estimates and Robustness to Differential Trends

	Dependent variable is log of population									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Interaction of post-year dummies with...									
	Institutions	Independent in 1940	Log GDP per capita in 1930	Log Population in 1930	Share Population 15-34 in 1940	Diamond Production per Capita in 1960	Oil and Gas Rents per Capita in 1960	Ethnic Polarization	Ethnic Fragmentation	Initial War in 1930s
Panel A: long differences, just 1940s and 1980s										
baseline predicted mortality	-0.730***	-0.815***	-0.600***	-0.762***	-0.747***	-0.776***	-0.808***	-0.599***	-0.587***	-0.935***
	(0.167)	(0.175)	(0.219)	(0.126)	(0.197)	(0.138)	(0.138)	(0.148)	(0.168)	(0.189)
Observations	102	102	100	102	86	102	102	96	96	88
R-squared	0.826	0.824	0.833	0.844	0.806	0.842	0.841	0.868	0.849	0.805
Number of clusters	50	50	49	50	43	50	50	47	47	44
p-value for post year dummy x variable indicated at the top of each column	0.461	0.582	0.0952	0.00980	0.321	0.000	0.000	0.00432	0.0605	0.931
Panel B: panel regressions, 1940s-1980s										
baseline predicted mortality	-0.437***	-0.484***	-0.361**	-0.446***	-0.440***	-0.460***	-0.478***	-0.337***	-0.330***	-0.559***
	(0.113)	(0.120)	(0.146)	(0.086)	(0.128)	(0.093)	(0.092)	(0.096)	(0.107)	(0.126)
Observations	307	300	267	265	223	300	300	277	277	244
R-squared	0.810	0.797	0.818	0.798	0.773	0.801	0.809	0.849	0.835	0.757
p-value for post year dummy x variable indicated at the top of each column	7.14e-05	0.00652	6.17e-10	0.0538	1.46e-08	0.000	0.000	0.000140	6.67e-05	0.0529
Number of clusters	63	61	53	52	45	61	61	56	56	50

Notes: OLS regressions with a full set of year and country fixed effects (equation (8) in the text). Robust standard errors (clustered by country) are reported in parentheses. Panel A presents long-difference specifications with two observations per country, one for the initial date and one for the final date. Panel B presents unbalanced panels with one observation per decade. For clustered standard errors, Bangladesh, India and Pakistan are considered a single cluster. See the text and Appendix Table A-1 for definitions and details.

Table 5: Reduced Forms and Falsification Exercises

	(1)	(2)	(3)	(4)	(5)	(6)
	Base Sample	Low and Middle Income Countries Only	Base Sample	Low and Middle Income Countries Only	Base Sample	Low and Middle Income Countries Only
	Reduced Form		Falsification Exercises			
Dependent Variable is:	Fraction of decade in conflict Change 1940s-1980s		Fraction of decade in conflict Change 1900s-1940s		Log Population Change 1900s-1940s	
Change in predicted mortality, 1940-1980	-0.491***	-0.660***	0.085	0.197	-0.189	-0.198
	(0.179)	(0.236)	(0.055)	(0.126)	(0.138)	(0.196)
Observations	52	41	36	28	52	41
R-squared	0.166	0.200	0.012	0.039	0.033	0.029

Notes: OLS regressions (equation (9) in the text). Robust standard errors (clustered by country) are reported in parentheses. For clustered standard errors, Bangladesh, India and Pakistan are considered a single cluster. See the text and Appendix Table A-1 for definitions and details.

Table 6: The Effect of Population on Conflict: 2SLS Estimates

	Dependent variable is fraction of decade in conflict (cols 1-3, 5-7), and Log of 1 + Battle Deaths/Pop. in 1940 (cols 4 and 8)							
	Baseline Predicted Mortality (cols 1-4)				Global Mortality Rate (cols 5-8)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	COW	Uppsala	Fearon & Laitin	Log(1+Battle Deaths/ Pop. 1940)	COW	Uppsala	Fearon & Laitin	Log(1+Battle Deaths/ Pop. 1940)
Panel A: long differences, just 1940s and 1980s								
log of population	0.617***	0.576**	0.879***	1.347**	0.624***	0.589**	0.872***	1.389**
	(0.213)	(0.238)	(0.303)	(0.598)	(0.216)	(0.247)	(0.303)	(0.619)
Observations	102	104	104	104	102	104	104	104
R-squared	0.056	0.064	-0.219	0.042	0.049	0.057	-0.212	0.033
Number of clusters	50	51	51	51	50	51	51	51
Panel B: panel regressions, 1940s-1980s								
log of population	0.570***	0.341	0.761**	1.278**	0.570***	0.341	0.761**	1.278**
	(0.191)	(0.233)	(0.334)	(0.501)	(0.191)	(0.233)	(0.334)	(0.501)
Observations	307	308	308	273	307	308	308	273
R-squared	0.014	0.145	0.001	0.074	0.014	0.145	0.001	0.074
Number of clusters	63	63	63	54	63	63	63	54

Notes: 2SLS regressions with a full set of year and country fixed effects (equation (4) in the text, where population is instrumented by predicted mortality, as in equation (7) in the text). Robust standard errors (clustered by country) are reported in parentheses. Panel A presents long-difference specifications with two observations per country, one for the initial date and one for the final date. Panel B presents unbalanced panels with one observation per decade. First stages, for the sample with data on years in conflict according to COW, in columns 1 and 4 of Table 3. For clustered standard errors, Bangladesh, India and Pakistan are considered a single cluster. See the text and Appendix Table A-1 for definitions and details.

Table 7: The Effect of Population on Conflict: 2SLS Estimates
Robustness to Differential Trends

	Dependent variable is fraction of decade in conflict according to Correlates of War									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
					Interaction of post-year dummies with...					
	Institutions	Independent in 1940	Log GDP per Capita in 1930	Log Population in 1930	Share Population 15-34 in 1940	Diamond Production per Capita in 1960	Oil and Gas Rents per Capita in 1960	Ethnic Polari-zation	Ethnic Fragmen-tation	Initial War in 1930s
Panel A: long differences, just 1940s and 1980s										
log of population	0.748***	0.563**	1.132***	0.617***	0.839***	0.623***	0.596***	0.738**	0.737**	0.497**
	(0.257)	(0.220)	(0.427)	(0.215)	(0.275)	(0.213)	(0.204)	(0.294)	(0.326)	(0.224)
Observations	102	102	100	102	86	102	102	96	96	88
p-value for post year dummy x variable indicated at the top of each column	0.106	0.515	0.0541	0.973	0.219	0.00626	0.0173	0.401	0.600	0.533
Number of clusters	50	50	49	50	43	50	50	47	47	44
Panel B: panel regressions, 1940s-1980s										
log of population	0.706***	0.497***	1.150**	0.617***	0.632**	0.614***	0.592***	0.712**	0.569**	0.319**
	(0.252)	(0.166)	(0.457)	(0.211)	(0.250)	(0.204)	(0.197)	(0.303)	(0.270)	(0.136)
Observations	307	300	267	265	223	300	300	277	277	220
p-value for post year dummies x variable indicated at the top of each column variable indicated at the top of each column	0.165	0.0707	0.0640	0.0160	0.0856	0.000	0.0958	0.562	0.111	0.00636
Number of clusters	63	61	53	52	45	61	61	56	56	44

Notes: 2SLS regressions with a full set of year and country fixed effects (equation (10) in the text, where population is instrumented by predicted mortality, as in equation (7) in the text). Robust standard errors (clustered by country) are reported in parentheses. Long-difference specifications with two observations per country, one for the initial date and one for the final date. First stages in Table 4. For clustered standard errors, Bangladesh, India and Pakistan are considered a single cluster. See the text and Appendix Table A-1 for definitions and details.

Table 8: The Effect of Population on Conflict: 2SLS Estimates
Basic Robustness

	Dependent variable is fraction of decade in conflict according to Correlates of War							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Base Sample	Excluding Eastern Europe	Low and Middle Income Countries Only	Global Mortality Instrument	Excluding Most Affected By WWII	Base Sample Assign 1950 to 1940	Base Sample Assign 1946-49 to 1940	Adding Population 15-34 as Covariate
Panel A: long differences, just 1940s and 1980s								
log of population	0.617***	0.657***	0.828***	0.624***	0.610***	0.420**	0.640***	0.704***
	(0.213)	(0.240)	(0.294)	(0.216)	(0.215)	(0.198)	(0.226)	(0.228)
Observations	102	92	80	102	94	102	102	86
Number of clusters	50	45	39	50	46	50	50	43
Panel B: panel regressions, 1940s-1980s								
log of population	0.609***	0.649***	0.862***	0.570***	0.615***	0.276**	0.649***	0.526**
	(0.205)	(0.241)	(0.307)	(0.191)	(0.211)	(0.110)	(0.236)	(0.212)
Observations	307	278	252	307	279	307	307	227
Number of clusters	63	57	52	63	57	63	63	46

Notes: 2SLS regressions with a full set of year and country fixed effects (equation (4) in the text, where population is instrumented by predicted mortality, as in equation (7) in the text). Robust standard errors (clustered by country) are reported in parentheses. Panel A presents long-difference specifications with two observations per country, one for the initial date and one for the final date. Panel B presents unbalanced panels with one observation per decade. First stages for columns 1-5, and 8 in columns 1-5 of Table 3. First stages of columns 6-7 are in column 1 of Table 3. For clustered standard errors, Bangladesh, India and Pakistan are considered a single cluster. See the text and Appendix Table A-1 for definitions and details.

Table 9: Timing of the Effect of Population on Conflict: 2SLS Estimates

	Dependent variable is fraction of decade in conflict according to Correlates of War				
	(1)	(2)	(3)	(4)	(5)
Panel A: Baseline Predicted Mortality					
Long differences	Just 1940s and 1960s	Just 1940s and 1970s	Just 1940s and 1980s	Just 1940s and 1990s	Just 1940s and 2000s
log of population	0.389	0.600**	0.617***	0.409**	0.296**
	(0.304)	(0.237)	(0.213)	(0.164)	(0.144)
Observations	102	102	102	102	102
Number of clusters	50	50	50	50	50
Panel B: Global Mortality					
Long differences	Just 1940s and 1960	Just 1940s and 1970s	Just 1940s and 1980s	Just 1940s and 1990s	Just 1940s and 2000s
log of population	0.461	0.574**	0.624***	0.384***	0.272**
	(0.345)	(0.229)	(0.216)	(0.147)	(0.131)
Observations	102	102	102	102	102
Number of clusters	50	50	50	50	50

Notes: 2SLS regressions with a full set of year and country fixed effects (equation (4) in the text, where population is instrumented by predicted mortality, as in equation (7) in the text). Robust standard errors (clustered by country) are reported in parentheses. Long-difference specifications with two observations per country, one for the initial date and one for the final date. First stages for columns Panel A in column 1 of of Table 3, and for Panel B in column 4 of Table 3. For clustered standard errors, Bangladesh, India and Pakistan are considered a single cluster. See the text and Appendix Table A-1 for definitions and details.

Table A-1: Variables and Sources

Variable	Description	Source
	Social Conflict	
Years in conflict/ Total years	Ratio of the number of years with an internal conflict to total years assigned to reference date. Assignment of years to reference dates and exact definition of internal conflict varies by data source, as detailed below.	
COW	Number of years with intra-state war (wars that predominantly take place within the recognized territory of a state). These wars include civil wars for central control (type 4 in the COW typology) or over local issues (type 5), as well as regional internal (type 6) and intercommunal (type 7) wars. Each war in the dataset may list more than one participating country. For example, the "Overthrow of Abd el-Aziz" involves Morocco and France, and the "First Lebanese" war involves Lebanon and the US. Despite French and American involvement, we take these as civil wars in Morocco and Lebanon, as fighting took place in their territory. The threshold for inclusion in the dataset is 1,000 battle-related deaths per year (twelve-month period beginning with the start date of the war) among all the qualified war participants, including deaths from combat wounds or from diseases contracted in the war theater. Assignment to reference dates: 1900=1900-09, 1940=1940-49, 1950=1950-59 ... 1990=90-1999, 2000=2000-2007. Downloaded from `http://www.correlatesofwar.org/datasets.htm`, last accessed on Sept 27, 2014.	Intra-State War Data (v4.0), Correlates of War (COW). Sarkees and Wayman (2010).
Uppsala	Number of years with any incidence of an "internal armed conflict" or of "internationalized internal armed conflict". Armed conflict is defined to include all contested incompatibilities that concern government or territory or both where the use of armed force between two parties results in at least 25 battle-related deaths. Of the two parties, at least one is the government of a state. Assignment to reference dates: 1940=1946-49, 1950=1950-59 ... 1990=1990-99, 2000=2000-2008. Downloaded with PRIO's Battle Deaths Dataset 3.0 (see below).	UCDP/PRIO Armed Conflict Dataset, Version 4-2008. Geldtisch et. al (2002).
Fearon and Laitin	Number of years with violent civil conflicts that: (1) involved fighting between agents of (or claimants to) a state and organized, nonstate groups who sought either to take control of a government, to take power in a region, or to use violence to change government policies, (2) killed at least 1,000 over its course, with a yearly average of at least 100, (3) At least 100 were killed on both sides (including civilians attacked by rebels). Counts anticolonial wars as occurring within the empire in question (e.g., Algeria is assigned to France). Assignment to reerence dates: : 1940=1945-49, 1950=1950-59 ... 1980=1980-89, 1990=1990-99.	Fearon and Laitin (2003)

Continued on next page

Table A-1 Variables and sources – continued from previous page

Variable	Description	Source
Log(1+Battle Deaths/ No. of years) and Log(1+Battle Deaths/ Pop. in 1940)	"Best estimate" of annual battle-related deaths for use with UCDP/PRIO dataset. Assignment to reference dates as in UCDP/PRIO dataset. Population in 1940 from Maddison (2003), see below. Downloaded from `http://www.prio.org/Data/Armed-Conflict/Battle-Deaths/The-Battle-Deaths-Dataset-version-30/`, last accessed September 19, 2014.	PRIO Battle Deaths Dataset version 3.0. Lacina and Gleditsch (2005)
	Population	
Log of Population	Total Population per country in 1900, 1940, 1950, 1960, 1970, 1980, 1990, 1990, 2000.	Maddison (2003)
Share of population 15-34	Percentage of the population ages 15-34 for 1940, 1950, 1960, 1970, 1980, 1990, 2000.	1950-1980: UN demographic database (`http://esa.un.org/unpp`). 1940: UN Demographic Yearbook 1948 (United Nations 1949, Table 4, pp. 108-158).
	Health	
Predicted Mortality Instrument	Sum of country's initial (in 1940) mortality rate from 15 diseases until there is a global intervention, and after the global intervention, the mortality rate from the disease in question declines to the frontier mortality rate. See paper for mathematical formula. 15 diseases are (in rough descending order of importance): malaria, pneumonia, and tuberculosis; influenza, cholera, typhoid, smallpox, shigella dysentery, whooping cough, measles (rubeola), dyphteria, scarlet fever, yellow fever, plague, typhus.	Acemoglu and Johnson (2007)
Global Mortality Instrument	Sum of the products of each country's initial (in 1940) mortality rate from 10 diseases and the ratio between the global mortality at time t to the initial (in 1940) global mortality from the disease in question. See paper for mathematical formula. Diseases are as for Predicted Mortality except yellow fever and dysentery/diarrhea for which it was not possible to track the diseases through changes in the classification of death over time. We also exclude cholera, typhoid, and plague since their were often not available for our extended sample of countries. Global Mortality is the unweighted average across countries in the sample of 59 countries (47 non-Eastern European countries and 12 countries with life expectancy data since 1950, see "Base Sample" below)	Acemoglu and Johnson (2007)
	Others	

Continued on next page

Table A-1 Variables and sources – continued from previous page

Variable	Description	Source
Base Sample	Includes Acemoglu and Johnson's (2007) list of 47 non-Eastern Europe countries (Argentina, Australia, Austria, Bangladesh, Belgium, Brazil, Canada, Chile, China, Colombia, Costa Rica, Denmark, Ecuador, El Salvador, Finland, France, Germany, Greece, Guatemala, Honduras, India, Indonesia, Ireland, Italy, South Korea, Rep., Malaysia, Mexico, Myanmar, Netherlands, New Zealand, Nicaragua, Norway, Pakistan, Panama, Paraguay, Peru, Philippines, Portugal, Spain, Sri Lanka, Sweden, Switzerland, Thailand, United Kingdom, United States, Uruguay, Venezuela); the set of 12 additional countries for which they have life expectancy data since 1950 (Algeria, Bolivia, Egypt, Iran, Iraq, Lebanon, Morocco, Singapore, South Africa, Tunisia, Turkey and Vietnam); and 6 countries from Eastern Europe (Bulgaria, Czech Republic, Hungary, Poland, Romania, and Russian Federation). This implies a total of 65 countries, but not all have all variables for all years. In particular, 13 countries lack population data and/or had not yet been created in 1940 (Algeria, Bangladesh, Egypt, Iran, Iraq, Lebanon, Malaysia, Morocco, Russia, Singapore, South Africa, Tunisia, and Vietnam). Also, Austria is excluded in 1940 when the dependent variables are from COW since it enters the COW state system in the 1950s.	
Initially rich, middle-income, and poor countries	Each category is defined using the top, middle, and lowest third group of countries in the base sample based on income in 1940. Initially rich countries had log GDP per capita over 8.4. in 1940; middle income had log GDP per capita betweeen 7.37 and 8.4; and low income countries had log GDP per capita below 7.37 in 1940.	
Country clusters	For clustered standard errors, Bangladesh, India and Pakistan are considered a single cluster.	
Countries most affected by World War II	Austria, China, Finland, Germany, Italy, Russian Federation	Acemoglu and Johnson (2007) based on Urlanis (2003)
	Controls and Baseline characteristics	
Institutions	Average of constraints on the executive in 1950, 1960 and 1970	Polity IV
Independent in 1940	=1 if country is independent in 1940, 0 otherwise	Own coding.
Initial GDP	Logarithm of GDP per capita in 1930	Maddison (2006)
Natural Resources	Oil Production and Oil Production per capita in 1960 Diamond Production and Diamond Production per capita in 1960 Share of natural resource sector in GNP in 1970 Share of mineral produciton in GNP in 1971 Oil and gas rents per capita in 1960	Humphreys (2005) Humphreys (2005) Sachs and Warner (1995) Sachs and Warner (1995) Ross (2008)
Ethnic and religious composition and polarization	Ethnic Polarization Ethnic Fragmentation Religious Polarization	Montalvo and Reynal-Querol (2005) Montalvo and Reynal-Querol (2005) Montalvo and Reynal-Querol (2005)

Continued on next page

Table A-1 Variables and sources – continued from previous page

Variable	Description	Source
	Religious Fragmentation	Montalvo and Reynal-Querol (2005)
	Ethnolinguistic fractionalization index (from 0 to 1). Average value of five indices based on ethnic or linguistic characteristics of the population.	Easterly and Levine (1997)
	Share of Muslim, Catholic and Protestant Populations in 1980.	La Porta et al (1999)

Tables below here are not for publication

Table A-2: Conflict and Predicted Mortality
Reduced Form Basic Robustness

	Dependent variable is fraction of decade in conflict according to Correlates of War							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Base Sample	Excluding Eastern Europe	Low and Middle Income Countries Only	Global Mortality Instrument	Excluding Most Affected By WWII	Base Sample Assign 1950 to 1940	Base Sample Assign centering 1946-49 to 1940	Adding Population 15-34 as Covariate
Panel A: long differences, just 1940s and 1980s								
baseline predicted mortality	-0.782***	-0.700***	-0.764***	-0.818***	-0.811***	-0.782***	-0.782***	-0.848***
	(0.141)	(0.141)	(0.191)	(0.145)	(0.140)	(0.141)	(0.141)	(0.173)
Observations	102	92	80	102	94	102	102	86
R-squared	0.823	0.847	0.828	0.823	0.842	0.823	0.823	0.832
Number of clusters	50	45	39	50	46	50	50	43
Panel B: panel regressions, 1940s-1980s								
baseline predicted mortality	-0.461***	-0.402***	-0.470***	-0.689***	-0.460***	-0.461***	-0.461***	-0.483***
	(0.094)	(0.093)	(0.129)	(0.128)	(0.096)	(0.094)	(0.094)	(0.118)
Observations	314	285	259	314	285	314	314	229
R-squared	0.794	0.820	0.816	0.817	0.819	0.794	0.794	0.756
Number of clusters	63	57	52	63	57	63	63	46

Notes: OLS regressions with a full set of year and country fixed effects (regressions as in equation (4) in the text, using predicted mortality instead of log population as a regressor). Robust standard errors (clustered by country) are reported in parentheses. Panel A presents long-difference specifications with two observations per country, one for the initial date and one for the final date. Panel B presents unbalanced panels with one observation per decade. For clustered standard errors, Bangladesh, India and Pakistan are considered a single cluster. See the text and Appendix Table A-1 for definitions and details.

Table A-3: Conflict and Predicted Mortality
Reduced Form Robustness to Differential Trends

	Dependent variable is fraction of decade in conflict according to Correlates of War									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Interaction of post-year dummies with...									
	Institutions	Independent in 1940	Log GDP per Capita in 1930	Log Population in 1930	Share Population 15-34 in 1940	Diamond Production per Capita in 1960	Oil and Gas Rents per Capita in 1960	Ethnic Polari-zation	Ethnic Fragmen-tation	Initial War in 1930s
Panel A: long differences, just 1940s and 1980s										
baseline predicted mortality	-0.730***	-0.815***	-0.600***	-0.762***	-0.747***	-0.776***	-0.808***	-0.599***	-0.587***	-0.935***
	(0.167)	(0.175)	(0.219)	(0.126)	(0.197)	(0.138)	(0.138)	(0.148)	(0.168)	(0.189)
Observations	102	102	100	102	86	102	102	96	96	88
R-squared	0.826	0.824	0.833	0.844	0.806	0.842	0.841	0.868	0.849	0.805
p-value for post year dummy x variable indicated at the top of each column	0.461	0.582	0.0952	0.00980	0.321	0	0	0.00432	0.0605	0.931
Number of clusters	50	50	49	50	43	50	50	47	47	44
Panel B: panel regressions,1940s-1980s										
baseline predicted mortality	-0.426***	-0.468***	-0.354**	-0.450***	-0.430***	-0.457***	-0.475***	-0.343***	-0.326***	-0.553***
	(0.111)	(0.114)	(0.139)	(0.086)	(0.126)	(0.093)	(0.092)	(0.091)	(0.101)	(0.126)
Observations	314	306	273	270	225	306	306	283	283	220
R-squared	0.812	0.799	0.816	0.799	0.773	0.803	0.811	0.851	0.837	0.754
p-value for post year dummies x variable indicated at top of each column	8.97e-05	0.0113	3.89e-10	0.0513	1.16e-08	0	0	0.000312	0.000472	0.0228
Number of clusters	63	61	53	52	45	61	61	56	56	44

Notes: OLS regressions with a full set of year and country fixed effects (regressions as in equation (10) in the text, using predicted mortality instead of (instrumented) log population as a regressor). Robust standard errors (clustered by country) are reported in parentheses. Long-difference specifications with two observations per country, one for the initial date and one for the final date. For clustered standard errors, Bangladesh, India and Pakistan are considered a single cluster. See the text and Appendix Table A-1 for definitions and details.

Table A-4: Predicted Mortality and Age Structure

	Dependent variable is share of population ages 15-34							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Baseline predicted mortality		Global mortality instrument		Baseline predicted mortality		Global mortality instrument	
	Base Sample	Low and Middle Income Countries Only	Base Sample	Low and Middle Income Countries Only	Base Sample	Low and Middle Income Countries Only	Base Sample	Low and Middle Income Countries Only
	Long differences, just 1940s and 1980s				Long differences, just 1940s and 1990s			
baseline predicted mortality	0.004	-0.000	0.005	0.001	-0.028	-0.030	-0.030	-0.032
	(0.017)	(0.028)	(0.018)	(0.031)	(0.019)	(0.035)	(0.021)	(0.039)
Observations	86	64	86	64	86	64	86	64
R-squared	0.017	0.025	0.017	0.026	0.057	0.038	0.057	0.038
Number of clusters	43	32	43	32	43	32	43	32

Notes: OLS regressions with a full set of year and country fixed effects (regressions as in equation (4) in the text, using share of population ages 15-34 instead of conflict as dependent variable). Robust standard errors (clustered by country) are reported in parentheses. For clustered standard errors, Bangladesh, India and Pakistan are considered a single cluster. See the text and Appendix Table A-1 for definitions and details.

Table A-5: Population and Predicted Mortality
First Stage Robustness to Additional Differential Trends

	Dependent variable is log of population							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			Interaction of post-year dummies with...					
	Share of GDP in Natural Resource Sector in 1970	Share of Mineral Production in GNP in 1971	Oil Production in 1960	Diamond Production in 1960	Religious Polarization	Religious Fragmentation	Average Ethno-linguistic Fragmentation	Share Catholic Muslim Protestant in 1980
Panel A: long differences, just 1940s and 1980s								
baseline predicted mortality	-0.495**	-0.661***	-0.796***	-0.772***	-0.562***	-0.636***	-0.777***	-0.584***
	(0.191)	(0.145)	(0.137)	(0.138)	(0.174)	(0.152)	(0.201)	(0.185)
Observations	68	92	102	102	96	96	102	100
R-squared	0.888	0.853	0.844	0.837	0.848	0.845	0.823	0.842
p-value for post year dummy x variable indicated at top of each column	0.953	0.167	1.63e-09	0.0126	0.0515	0.0588	0.961	0.158
Number of clusters	33	45	50	50	47	47	50	49
Panel B: panel regressions, 1940s-1980s								
baseline predicted mortality	-0.268**	-0.379***	-0.469***	-0.455***	-0.312***	-0.362***	-0.429***	-0.359***
	(0.116)	(0.095)	(0.092)	(0.092)	(0.110)	(0.098)	(0.119)	(0.123)
Observations	197	273	306	306	283	283	298	301
R-squared	0.871	0.835	0.813	0.801	0.839	0.829	0.804	0.834
p-value for post year dummies x variable indicated at top of each column	8.97e-06	1.97e-05	0	7.52e-06	1.57e-07	0.000311	0.000516	2.44e-10
Number of clusters	38	54	61	61	56	56	59	60

Notes: OLS regressions with a full set of year and country fixed effects (equation (8) in the text). Robust standard errors (clustered by country) are reported in parentheses. Long-difference specifications with two observations per country, one for the initial date and one for the final date. For clustered standard errors, Bangladesh, India and Pakistan are considered a single cluster. See the text and Appendix Table A-1 for definitions and details.

Table A-6: Conflict and Population
2SLS Estimates Robustness to Additional Differential Trends

	Dependent variable is fraction of decade in conflict according to Correlates of War							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			Interaction of post-year dummies with...					
	Share of GDP in Natural Resource Sector in 1970	Share of Mineral Production in GNP in 1971	Oil Production in 1960	Diamond Production in 1960	Religious Polarization	Religious Fragmen-tation	Average Ethno-linguistic Fragmen-tation	Share Catholic Muslim Protestant in 1980
Panel A: long differences, just 1940s and 1980s								
log of population	1.415**	0.720***	0.605***	0.627***	0.848**	0.772***	0.555**	0.965**
	(0.558)	(0.265)	(0.206)	(0.214)	(0.371)	(0.288)	(0.275)	(0.380)
Observations	68	92	102	102	96	96	102	100
p-value for post year dummy x variable indicated at top of each column	0.0553	0.0632	0.00768	0.0358	0.365	0.270	0.678	0.161
Number of clusters	33	45	50	50	47	47	50	49
Panel B: panel regressions, 1940s-1980s								
log of population	1.498**	0.667***	0.600***	0.618***	0.710**	0.672**	0.356**	0.945**
	(0.668)	(0.259)	(0.203)	(0.205)	(0.353)	(0.275)	(0.181)	(0.410)
Observations	191	267	300	300	277	277	292	295
p-value for post year dummies x variable indicated at top of each column	0.165	0.0166	0.0804	6.62e-11	0.701	0.472	0.147	0.441
Number of clusters	38	54	61	61	56	56	59	60

Notes: 2SLS regressions with a full set of year and country fixed effects (equation (10) in the text). Robust standard errors (clustered by country) are reported in parentheses. Long-difference specifications with two observations per country, one for the initial date and one for the final date. For clustered standard errors, Bangladesh, India and Pakistan are considered a single cluster. First stages in Table A-5. See the text and Appendix Table A-1 for definitions and details.

Table A-7: Conflict and Predicted Mortality
Reduced Form Robustness to Additional Differential Trends

	Dependent variable is fraction of decade in conflict according to Correlates of War							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			Interaction of post-year dummies with...					
	Share of GDP in Natural Resource Sector in 1970	Share of Mineral Production in GNP in 1971	Oil Production in 1960	Diamond Production in 1960	Religious Polarization	Religious Fragmentation	Average Ethno-linguistic Fragmentation	Share Catholic Muslim Protestant in 1980
Panel A: long differences, just 1940s and 1980s								
baseline predicted mortality	-0.700***	-0.476**	-0.481***	-0.484***	-0.477**	-0.492**	-0.431**	-0.563***
	(0.232)	(0.182)	(0.178)	(0.179)	(0.228)	(0.201)	(0.199)	(0.208)
Observations	68	92	102	102	96	96	102	100
R-squared	0.331	0.210	0.195	0.196	0.194	0.194	0.199	0.265
p-value for post year dummy x variable indicated at top of each column	0.0319	0.331	0.551	0.115	0.942	0.924	0.660	0.212
Number of clusters	33	45	50	50	47	47	50	49
Panel B: panel regressions, 1940s-1980s								
baseline predicted mortality	-0.370***	-0.247***	-0.285***	-0.292***	-0.204*	-0.231**	-0.185***	-0.318***
	(0.131)	(0.084)	(0.086)	(0.088)	(0.107)	(0.099)	(0.068)	(0.114)
Observations	192	271	305	305	281	281	295	300
R-squared	0.116	0.094	0.077	0.078	0.090	0.086	0.097	0.136
p-value for post year dummies x variable indicated at top of each column	0.397	0.365	0.832	8.41e-10	0.660	0.633	0.0599	0.0460
Number of clusters	38	54	61	61	56	56	59	60

Notes: OLS regressions with a full set of year and country fixed effects (regressions as in equation (10) in the text, using predicted mortality instead of (instrumented) log population as a regressor). Robust standard errors (clustered by country) are reported in parentheses. Long-difference specifications with two observations per country, one for the initial date and one for the final date. For clustered standard errors, Bangladesh, India and Pakistan are considered a single cluster. See the text and Appendix Table A-1 for definitions and details.