

# Population and Conflict

DARON ACEMOGLU

*MIT*

LEOPOLDO FERGUSSON

*Universidad de los Andes*

and

SIMON JOHNSON

*MIT*

*First version received April 2017; Editorial decision July 2019; Accepted August 2019 (Eds.)*

Medical innovations during the 1940s quickly resulted in significant health improvements around the world. Countries with initially higher mortality from infectious diseases experienced larger increases in life expectancy, population, and subsequent social conflict. This cross-country result is robust across alternative measures of conflict and is not driven by differential trends between countries with varying baseline characteristics. A similar effect is also present within Mexico. Initial suitability conditions for malaria varied across municipalities, and anti-malaria campaigns had differential effects on population growth and social conflict. Both across countries and within Mexico, increased conflict over scarce resources predominates and this effect is more pronounced during times of economic hardship (specifically, in countries with a poor growth record and in drought-stricken areas in Mexico). At least during this time period, a larger increase in population made social conflict more likely.

*Key words:* Civil war, Conflict, Mortality, Natural resources, Population growth

*JEL Codes:* J1, O11, O15, Q56

## 1. INTRODUCTION

The world's population is forecast to rise from its current level of around 7.4 billion to nearly 11.2 billion by 2100. This growth will be unequally distributed around the world. "More developed regions," as classified by the UN, are expected to remain at roughly the same population level, 1.25–1.30 billion, over the next century. "Least developed countries," on the other hand, are projected to increase their population from 957 million in 2015 to 3.2 billion in 2100. The UN forecasts that the population of Africa will rise from its current level of near 1.2 billion to almost 4.5 billion within the century.<sup>1</sup> What is the likely impact of a population increase on this scale?

1. The data in this paragraph are from the UN's 2017 long-term population projections, downloaded on 15 December 2018 from <https://population.un.org/wpp> (United Nations, 2015). Latest annual data are for 2015 and the

---

*The editor in charge of this paper was Nicola Gennaioli.*

These population changes are not entirely unprecedented. During the *international epidemiological transition* starting in the 1940s, largely due to the introduction of new drugs, chemicals, and public health measures, many relatively less prosperous countries experienced major improvements in health and longevity (Acemoglu and Johnson, 2007). These innovations led to large increases in population. For example, between 1940 and 1980, the population of the poorest third of countries in our sample increased from around 1.13 billion to 2.55 billion.<sup>2</sup>

In this article, we exploit this historical episode to shed light on whether changes in population have a major impact on civil wars and conflict. We focus on the population increase between 1940 and 1980, during which global health technology and conditions improved dramatically.

In our cross-country analysis, we exploit variation in population growth coming from the initial distribution of mortality from infectious 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 and likelihood of internal conflict in the rest of the world. Our identification strategy does not depend on when a particular country adopted better public health measures or how effectively these measures were implemented. We also document that there were no differential trends in life expectancy, population, conflict, or economic outcomes between initial high and low mortality countries before the onset of the international epidemiological transition.

Countries with higher exogenous increases in population experienced more social conflict after 1940. Across different specifications and alternative definitions of conflict, instrumented changes in population have a significant positive effect on various measures of civil war and violent conflict.

The effect of population increase on social conflict is sizable. A rise in log population of about 0.68 from 1940 to 1980, corresponding to the average change in population in our sample of countries, caused roughly 4.2 additional years of full-blown civil war in the 1980s relative to the 1940s.

An important concern is that the effects of changes in mortality from various diseases on social conflict might work through channels other than population. For example, these mortality changes also impact the overall health of the population and the age structure and may lead to greater education for new cohorts, all of which potentially affect conflict. Although we explore the various channels and show that they do not appear to have a first-order impact on our measures of conflict and cannot account for our results, this concern implies that our instrumental-variables results ought to be interpreted with caution.

Our preferred interpretation for the linkages between disease mortality, population, and conflict is via a Malthusian channel: growth in population unaccompanied by improved productivity or economic opportunities intensifies conflict over scarce resources. We provide a number of results consistent with this channel, in particular showing that population surges increase conflict related to natural resources but have no effect on conflicts unrelated to natural resources, and that they raise conflict in slowly growing—but not in rapidly growing—countries.

projections run through 2100. According to the notes provided by the UN with the downloaded data: “More developed regions comprise Europe, Northern America, Australia/New Zealand, and Japan,” and “The group of least developed countries, as defined by the United Nations General Assembly included 48 countries in January 2014: 34 in Africa, 9 in Asia, 4 in Oceania, and one in Latin America and the Caribbean.”

2. Because our study requires detailed death-by-disease data, our base sample contains some—but not all—countries that were low income in 1940. Our sample has good coverage for 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, since most sub-Saharan African countries lack reliable data on causes of death disaggregated by disease dating back to the 1940s, and these data are essential for our empirical strategy.

The second part of the article examines the experience in Mexico. In 1940, Mexico had an income per capita about 25% of the U.S. level, slightly above the more prosperous parts of Asia and about three times the income level of India and many parts of sub-Saharan Africa. Most important for our purposes, Mexican municipalities experienced differential changes in mortality from malaria depending on their initial exposure (as in our cross-country design). This differential incidence of malaria within Mexico was due almost entirely to differences in climate and geography. Starting in the 1940s, as part of the broader international epidemiological transition, Mexico embarked upon an ambitious anti-malarial campaign. These efforts quickly brought down morbidity and mortality rates from malaria.

We find that increases in population across Mexican municipalities driven by the successful anti-malaria campaign are associated with more conflict, in particular, with more violent protests during the 1960s. Consistent with our Malthusian interpretation, the significant increases in violent protests are driven by those related to natural resources. Moreover, these protests were more pronounced when there was also economic hardship in the form of drought during the growing season.

Our article is related to several strands of research. There is an extensive literature on malaria and the effectiveness of anti-malaria campaigns from a health perspective. Prominent papers include Barlow (1967), Case and Paxson (2009), Hong (2007), Ashraf *et al.* (2010), Lucas (2010), Flückiger and Ludwig (2017), and Barofsky *et al.* (2015). This literature does not examine the potential impact of population and health on social conflict.

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. Most of this literature does not fully address the possibility that reverse causality or omitted variables bias may be driving the observed correlations. Blattman and Miguel (2010) conclude 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). Miguel *et al.* (2004) use annual rainfall as an instrument for income growth in sub-Saharan Africa, while Besley and Persson (2008) rely on plausibly exogenous international commodity price movements.

Although population has not been a prime focus in the economics of conflict literature (see the survey by Garfinkel and Skaperdas, 2007), there has been a lively debate on the effects of population pressure on violent conflict in other disciplines, including political science. For example, Homer-Dixon (1991, 1999) studies the connection between population growth, pressure on environmental resources and conflict, and finds that poor countries are in general more vulnerable to environmentally induced conflicts. This view motivated Robert Kaplan’s famous 1994 essay, “The Coming of Anarchy”, which predicted mounting conflict around the world. Similar findings and the resulting stress on state’s capacity to manage conflicts have been documented in Jack Goldstone’s work (see, for instance, Goldstone 1991, 2002; Goldstone *et al.* 2012). However, the literature is far from a consensus, and several authors, including Richards (1996), have pushed back against this view. Consistent with the notion that population pressures contribute to conflict, Tir and Diehl (1998) estimate that population growth pressures have a significant impact on some dimensions of military conflict, especially in poor countries. Hauge and Ellingsen (1998) find that factors like deforestation, land degradation, and water scarcity, alone and in combination with high population density, increase the risk of domestic (low-level) armed conflict. Urdal (2005) does not find a strong correlation between population growth and conflict risk, but confirms that conflict is more likely when high population growth combines with land scarcity (see also Diehl and Gleditsch, 2001).

To the best of our knowledge, only Brückner (2010) attempts to establish the causal impact of population size on conflict. He exploits randomly occurring droughts as an instrument for population to address endogeneity and finds that a 5% increase in population size raises the risk of civil conflict by around six percentage points. However, Brückner's study focuses on 37 sub-Saharan countries between 1981 and 2004, where the effects of population pressures or droughts may be different than in other settings.

In Section 2, we present a simple motivating model where a Malthusian mechanism links population to conflict. Section 3 describes the data. Section 4 presents our cross-country analysis, including extensive robustness checks. Section 5 presents our within-country analysis for Mexico. Section 6 concludes. Additional robustness checks and details are provided in the Supplementary Appendix.

## 2. A MODEL OF MALTHUSIAN CONFLICT

In this section, we present a simple framework capturing the Malthusian notion that population growth may lead to social conflict (Malthus, 1798). The basic idea is that higher population generates greater rents for fixed factors relative to labour, and this scarcity makes conflict more likely. For less developed economies in 1940 or today, it makes sense to think of land and natural resources as the scarce factor (henceforth, just land).

An important conclusion from this framework is that not every increase in population leads to conflict. Only population growth that intensifies the scarcity of land intensifies conflict. Conversely, population growth associated with improvements in technology or institutions will not promote greater conflict. Nor will population growth in economies where scarce factors are unimportant.

Suppose that aggregate output is given by a constant returns to scale (differentiable) production function where the factors of production are land (in fixed supply) denoted by  $Z$  and labour,  $N$ . In addition, the production function is indexed by technology,  $A$ :

$$Y = F(Z, N, A) \equiv f(N). \quad (2.1)$$

Here  $F(\cdot)$  exhibits constant returns to scale in  $(Z, N)$  and  $f$ , which gives output as a function of labour, holding technology,  $A$ , and land,  $Z$ , constant, has diminishing returns. Consequently, when  $N$  increases and  $A$  and  $Z$  remain constant, the fixed factor of production,  $Z$ , will become more scarce, and so output per worker,  $f(N)/N$  and the marginal product of labour,  $f'(N)$  will decline. On the other hand, when increases in  $N$  are accompanied by improving technology, output per worker and wage need not decline.

We assume the following simple allocation of resources. Each individual  $i$  in society supplies one unit of labour inelastically and also owns a fraction  $\theta_i$  of land. For simplicity, we suppose that all markets are competitive, though this is not important for our analysis. With these assumptions, individual income and consumption (without conflict) are given by

$$c_i(N, \theta_i) = f'(N) + \theta_i [f(N) - Nf'(N)]. \quad (2.2)$$

The key observation from equation (2.2) is that a marginal increase in an individual's consumption from an increase in his landholdings is larger when population is greater, *i.e.*

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

Put differently, land shares matter more for consumption when population is greater. The intuition is straightforward: with higher  $N$ , land rents are more important relative to wages due to

the diminishing marginal product of labour. This is the fundamental force underpinning our *Malthusian model of conflict*, which implies that the greater scarcity increases conflict over the scarce factor.

To explore this channel in greater detail, imagine society consists of two groups, 1 and 2. All members within a group are identical. To simplify the discussion suppose that both groups are of size  $N/2$  and population growth leaves relative shares unchanged. Let us also assume that during conflict each group loses a fraction  $\rho$  of its labour income.

Group  $j$  has probability  $p_j$  of winning the conflict and if it does win, it captures a fraction  $\lambda_{-j}$  of the land of the other group, where  $\lambda$  is 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  $\lambda_j$  of its land. Also for simplicity, we ignore any first-mover advantage, and there are no deaths from any conflict. Moreover, as we elaborate further below, for now we do not allow any voluntary transfers or concessions to avoid conflict. Finally, we assume that all agents are risk neutral. Then the expected benefits from conflict,  $\pi_j(N, \theta, \lambda, \rho)$ , for group  $j$  are given by

$$\pi_j(N, \theta, \lambda, \rho) = -\rho f'(N) + [p_j \lambda_{-j} \theta_{-j} - p_{-j} \lambda_j \theta_j] [f(N) - Nf'(N)]. \tag{2.3}$$

The first term of this expression represents the loss of labour income. Because of this cost, no group may wish to initiate conflict. If there is going to be any conflict, it will be the group, here group  $j$ , for which

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

that has an incentive to initiate conflict.

The same reasoning as in our discussion of equation (2.2) implies that for the group at the margin of initiating conflict, we have:<sup>3</sup>

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

Therefore, an increase in population makes the group that is more likely to initiate conflict more “pro-conflict”. As noted before, this result does not apply when  $A$  increases in tandem with  $N$ .<sup>4</sup> This highlights 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 observation underpins one of the implications we investigate in our empirical work: population surges should have a bigger impact on conflict in economies that are already under pressure from either slow growth or other economic problems.

This simple framework generates some intuitive comparative static results. Suppose for specificity that group  $j$  is at the margin of initiating conflict. Then, a greater share of resources accruing to the weaker group ( $\theta_j$ ) makes conflict more likely. Lower disruption costs (lower  $\rho$ ) and lower asset specificity of the weaker group (higher  $\lambda_{-j}$ ) also increase the likelihood of conflict.<sup>5</sup>

3. In particular,  $\frac{\partial \pi_j(N, \theta, \lambda, \rho)}{\partial N} = -\rho f''(N) + [p_j \lambda_{-j} \theta_{-j} - p_{-j} \lambda_j \theta_j] [-Nf''(N)] > 0$  since  $f'' < 0$  and for the group at the margin of initiating conflict  $p_j \lambda_{-j} \theta_{-j} - p_{-j} \lambda_j \theta_j > 0$ .

4. A change that increases  $f'(N)$  and  $f(N) - Nf'(N)$  by the same proportional amount would not change incentives for conflict. More explicitly, starting from  $\pi_j(N, \theta, \lambda, \rho) = 0$ , a change that increases both  $f'(N)$  and  $f(N) - Nf'(N)$  by the same proportion will leave  $\pi_j(N, \theta, \lambda, \rho) = 0$ .

5. The relationship between inequality and conflict is non-monotonic. If  $\theta_j = \theta_{-j} = 1/2$ , an increase in  $\theta_{-j}$  increases both inequality and the likelihood of conflict. But if  $\theta_{-j} < \theta_j$ , then an increase in  $\theta_{-j}$  reduces economic inequality but still 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. A more industrial economy heavily relies on production processes—such as factories and long supply chains—that can be easily disrupted with violence, and skilled human capital, which is typically harder to expropriate. This increases specificity and discourages conflict (Acemoglu and Robinson, 2006). In contrast, the costs of conflict are smaller in economies relying on agriculture and natural resources. This reasoning highlights another implication we investigate empirically: population growth should lead to conflict over scarce natural resources and not necessarily increase conflict unrelated to resources.

Finally, a central question that we have so far ignored is why more efficient ways of redistributing resources do not prevent conflict. A plausible explanation concerns commitment problems (Fearon, 1998, 2004; Acemoglu and Robinson, 2001, 2006; Powell, 2006; Acemoglu *et al.*, 2012). 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 conflict. Assume all agents have a discount factor  $\beta \in (0, 1)$ . To simplify the discussion, assume as well that after conflict there is a permanent redistribution of resources and never any social conflict again, and that only cash transfers (and no asset transfers) are feasible. Suppose again that it is group  $j$  that is considering to initiate a conflict. In this context, the benefits from conflict for group  $j$  are proportional to  $1/(1-\beta)$  because of discounting. If the group is sufficiently patient ( $\beta$  is high enough), then cash transfers in a given period are not sufficient to offset this gain. But group  $-j$  cannot make a credible promise to make the cash transfers in the future once the window of opportunity for initiating conflict disappears. In this setting, there will be conflicts 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 conflict. Also, there exists  $\hat{q} < \bar{q}$ , so that for all  $q < \hat{q}$ , all Subgame Perfect Equilibria involve conflict (see Acemoglu and Robinson, 2006). It is straightforward to see that in this dynamic extension with voluntary transfers it is still the Malthusian channel that is responsible for conflict, and thus the comparative statics we have highlighted will apply—greater scarcity will make conflict more likely.

### 3. DATA

In this section, we present our cross-country and within-Mexico data and sources.

#### 3.1. Cross-country data

In our baseline analysis, we measure conflict as the ratio of the number of years in conflict to total years in a period. We refer to these periods with the beginning year (*e.g.* we refer to the average number of years in conflict between 1940 and 1950 as conflict in 1940).<sup>6</sup> This measure captures general conflict incidence rather than the precise timing of conflict. In our context, 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. Datasets sometimes disagree on the exact year when a conflict began, but there are fewer differences regarding the incidence of conflict within a decade.

Our baseline dataset is version 4 of the Correlates of War (henceforth COW) dataset (Sarkees and Wayman, 2010). 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

6. After 1940, some countries became independent, others lost their independence, fragmented, or experienced a major changes 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.



actively involved in military action, with effective resistance from both sides and with at least 1,000 battle-related deaths during the war. 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 1940–9 are assigned to 1940, wars from 1950–9 are assigned to 1950, and so on.

Our second database, covering dates since 1946, is the Uppsala Conflict Data Project, in conjunction with International Peace Research Institute (UCDP/PRIO Armed Conflict Dataset Version 4, [Gleditsch et al., 2002](#)). We assign the number of years in conflict to reference dates as follows: 1946–9 to 1940, 1950–9 to 1950, 1960–9 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 1946); for other reference years we divide it 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 1945–99, and the criteria are broadly similar to those of COW, 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 corresponds to 1945–9, 1950 corresponds to 1950–9, etc.).<sup>7</sup>

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 to potential mechanical effects; for example, 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 = 1940–9, 1950 = 1950–60, etc.)

We have partial data for the 65 countries listed in Supplementary Appendix Table A-1 (see “Base Sample”) and complete data from 1940 or earlier for 52 countries (51 when using COW since Austria enters the COW state system in the 1950s). Unfortunately, there are no reliable historical data on causes of death for sub-Saharan Africa during the period under investigation.

We also consider a number of control variables in our robustness exercises, all of which are described in Supplementary Appendix Table A-1.

### 3.2. Mexican data

The unit of analysis for Mexico is a municipality, of which there are about 2,400, covering the entire country. Population growth and other municipal characteristics—age composition, school and university enrolment, literacy, and share of immigrants—come from historical national censuses, accessed through the Mexican National Statistical Agency (INEGI). We use population data from the Archivo Histórico de Localidades (AHL), compiled by the INEGI, for localities

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

(the level of territorial division below municipalities). Municipality population is then computed as the sum of the populations of the relevant localities. We adopt [Sellars and Alix-Garcia's \(2018\)](#) corrections for missing localities in the AHL data.

We also use data from [Matsuura and Willmott \(2009\)](#) to code droughts in the 1960s, which we define as the number of months for which precipitation is below the 5th percentile of the long-run distribution (1900–2008) of monthly rain per municipality.

Our measure of malaria suitability is an index of potential malaria (plasmodium vivax) transmission based on temperature suitability during an average year ([Gething et al., 2011](#)). The index is available for 1 km by 1 km cells. Values range from 0 (completely unsuitable for transmission) to 1 (maximum suitability). We average this index across the cells in the municipality.

To measure conflict at the municipality level, we use information on social conflict in the 1960s collected by [Fergusson et al. \(2018\)](#) from *Excelsior* and *El Universal*, the only two independent newspapers with national coverage. From hand-coded news stories on protests, strikes, demonstrations, riots, and marches using the universe of articles published in the 1960s, we construct counts of violent and non-violent protests. More specifically, we used all news stories about protests published from 1960 to 1969 in these two newspapers. We first identified any news story including any of the following keywords in the title, description, or main text: *Protestas* (protests) and the n-gram “protest\*”, *Huelgas* (strikes) and the n-gram “huelg\*”, *Manifestaciones* (demonstrations) and the n-gram “manifesta\*”, *Disturbios* (riots) and the n-gram “Disturbio\*”, *Marchas* (marches) and the n-gram “March\*”. We then identified stories in which the title or description is related to violence, conflict, arms, social disorder, or aggression. Specifically, we again use keywords to code news stories as related to violent protests: “agita”, “desorden”, “violencia”, “violación”, “armado”, “agresión”, and “conflicto”. To measure non-violent protests, we simply counted all news stories about protests that do not have these keywords for violence. We also use keywords to identify protests related to resources and those that are unrelated to resources. All of our measures are expressed in per capita terms relative to 100,000 people in 1940. Prior to the 1960s, the ruling party (Partido Revolucionario Institucional, PRI) had a strong grip on the country and on the media, so there was little or no protest. In contrast, during the 1960s, some parts of the country became much more prone to protest, including violent conflicts in some instances, and our data confirm this.<sup>8</sup>

In robustness checks, we use data on historical conflict compiled by [Ramos-Toro \(2018\)](#).<sup>9</sup> Additional variables used in our robustness exercises are described in Supplementary Appendix Table A-1.

#### 4. CROSS-COUNTRY EVIDENCE

In this section, we present our cross-country evidence. Our focus is on two stage least squares (2SLS) estimates exploiting differences in predicted mortality. Nevertheless, for comparison we start with descriptive statistics and Ordinary Least Squares (OLS) estimates.

8. Specifically: the student movement became important; peasants defied the National Peasant Union (CNC), the PRI's arm in the rural sector; and opposition parties started to challenge the PRI in some parts of the country. The PRI stayed in power both nationally and locally with a combination of fraud and repression ([Bezdek, 1973](#); [Bartra, 1985](#); [Bellingeri, 2003](#); [Herrera Calderón and Cedillo, 2012](#)).

9. [Fergusson et al. \(2018\)](#) show that these protest counts are strongly correlated with later electoral support for political opposition, once reliable voting data become available for Mexico in the 1980s.



TABLE 1  
*Descriptive statistics*

Variable	Year	Base sample	By initial income			By declines in predicted mortality	
			Rich	Middle	Poor	Above median	Below median
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Fraction of decade in conflict, COW	1900	0.029 (0.099)	0 (0)	0.050 (0.134)	0.011 (0.033)	0.047 (0.131)	0.006 (0.025)
Fraction of decade in conflict, COW	1940	0.072 (0.212)	0 (0)	0.094 (0.222)	0.087 (0.263)	0.098 (0.261)	0.044 (0.142)
Fraction of decade in conflict, COW	1980	0.118 (0.260)	0 (0)	0.079 (0.188)	0.216 (0.347)	0.209 (0.328)	0.025 (0.102)
Fraction of decade in conflict, UCDP/PRIO	1940	0.126 (0.306)	0 (0)	0.115 (0.276)	0.229 (0.417)	0.201 (0.369)	0.048 (0.200)
Fraction of decade in conflict, UCDP/PRIO	1980	0.269 (0.408)	0.091 (0.302)	0.190 (0.341)	0.440 (0.468)	0.400 (0.447)	0.134 (0.319)
Fraction of decade in conflict, Fearon-Laitin	1940	0.121 (0.302)	0 (0)	0.100 (0.255)	0.237 (0.427)	0.163 (0.347)	0.077 (0.247)
Fraction of decade in conflict, Fearon-Laitin	1980	0.242 (0.413)	0.091 (0.302)	0.183 (0.369)	0.376 (0.475)	0.367 (0.467)	0.113 (0.306)
log (1 + battle deaths/population)	1940	0.157 (0.527)	0 (0)	0.251 (0.707)	0.112 (0.305)	0.208 (0.644)	0.105 (0.373)
log (1 + battle deaths/population)	1980	0.282 (0.779)	0.002 (0.005)	0.159 (0.528)	0.638 (1.153)	0.541 (1.024)	0.005 (0.017)
Fraction of decade in conflict, natural-resource	1940	0.019 (0.137)	0 (0)	0 (0)	0.062 (0.250)	0.037 (0.192)	0 (0)
Fraction of decade in conflict, natural-resource	1980	0.197 (0.397)	0 (0)	0.100 (0.300)	0.396 (0.495)	0.327 (0.471)	0.062 (0.246)
Fraction of decade in conflict, non-resource	1940	0.135 (0.309)	0 (0)	0.115 (0.276)	0.260 (0.417)	0.182 (0.339)	0.087 (0.273)
Fraction of decade in conflict, non-resource	1980	0.125 (0.281)	0.091 (0.302)	0.103 (0.265)	0.164 (0.296)	0.155 (0.305)	0.094 (0.254)
log population	1900	8.610 (1.617)	8.922 (1.410)	8.238 (1.394)	8.952 (2.016)	8.299 (1.668)	8.909 (1.541)
log population	1940	9.136 (1.455)	9.349 (1.344)	8.738 (1.192)	9.557 (1.756)	9.010 (1.530)	9.261 (1.393)
log population	1980	9.812 (1.384)	9.762 (1.293)	9.393 (1.238)	10.321 (1.461)	9.856 (1.484)	9.768 (1.294)
Baseline predicted mortality	1940	0.469 (0.271)	0.171 (0.050)	0.487 (0.224)	0.626 (0.272)	0.690 (0.195)	0.241 (0.080)

*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 UCDP/PRIO sources in the 1940s. See the text and Supplementary Appendix Table A-1 for details and definitions.

#### 4.1. Descriptive statistics

Table 1 presents descriptive statistics (sample means and standard deviations) for our baseline cross-country sample. We present these summary statistics for our entire sample, and for subsamples of countries split by income and by whether they have experienced declines in predicted mortality above or below the median. Countries above median predicted mortality are those expected to experience greater “exogenous” declines in mortality and increases in population.

The first two rows show an increase in conflict in our sample between 1900 and 1940. However, columns 6 and 7 show that this increase is not correlated with the later changes in predicted mortality. The increase is larger in absolute terms for the “above-median” sample (0.051 versus 0.038), but since the group of countries with below-median changes in predicted mortality starts off at a lower baseline incidence (0.006 versus 0.047), the relative increase is marginally larger in the “below-median” group. The next six rows of column 2 show a general trend, evident across all measures, of increasing conflict from the 1940s to the 1980s. Columns 3–5 show that this increase is concentrated among middle-income and especially among poor countries. 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 rose from 0.98 years in 1940 to 2.09 years in 1980 for countries with above median change in predicted mortality, while it decreased from 0.44 years to 0.25 years for those with below-median change. Table 1 further shows that this increase is accounted for by conflicts over natural resources, with no evidence for a differential change in non-resource conflicts. Our regression analysis confirms these patterns and establishes their robustness.

Finally, we also see a larger increase in population for the above-median group between 1940 and 1980, and no differential changes in population between 1900 and 1940 (see rows 14–16).

#### 4.2. OLS results

We begin with simple OLS regressions of conflict on population. In Table 2, we report regressions of the form

$$c_{it} = \pi x_{it} + \zeta_i + \mu_t + \mathbf{Z}'_{it} \beta + \varepsilon_{it}, \quad (4.4)$$

where  $c_{it}$  is a measure of conflict for country  $i$  and reference year  $t$ , and  $x_{it}$  is log population.  $\zeta_i$  denotes a full set of country fixed effects while  $\mu_t$  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.  $\mathbf{Z}_{it}$  is a vector of other controls. For all of our regressions, we calculate standard errors that are robust against heteroscedasticity and serial correlation at the country level (e.g. as in Wooldridge, 2002, p. 275).<sup>10</sup>

In Table 2 and hereafter, we present two types of estimates: 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. conflict in the 1980s, assigned to 1980). In these specifications, equation (4.4) 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 because of data availability. In Section 4.6, we investigate how the response of conflict to population growth changed over time.

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 in the long-difference regression in column 1 of Panel A, 0.323, implies that a 10% increase in population will be associated with a

10. One concern is that these standard errors may be downward biased due to a small number of clusters. We also implemented the wild bootstrap procedure suggested by Cameron *et al.* (2008). These results are presented in columns 1 and 2 of Supplementary Appendix Table A-2 and confirm that our conclusions are not affected by the exact method of computing standard errors. This is consistent with Cameron *et al.* (2008) who find very similar rejection rates for the cluster robust and wild bootstrap standard errors in their Monte Carlo simulations with 50 clusters.

TABLE 2  
*Population and conflict: OLS estimates*

Dependent variable...	Fraction of decade in conflict			log(1+battle deaths/ pop. 1940)
	COW	UCDP/PRIO	Fearon and Laitin	
	(1)	(2)	(3)	(4)
Panel A: Long differences, just 1940s and 1980s				
log population	0.323 (0.116)	0.271 (0.135)	0.236 (0.139)	0.722 (0.401)
Observations	102	104	104	104
$R^2$	0.177	0.145	0.098	0.102
Number of clusters	50	51	51	51
Panel B: Panel regressions, 1940s–1980s				
log population	0.268 (0.095)	0.311 (0.132)	0.251 (0.132)	0.738 (0.375)
Observations	307	308	308	273
$R^2$	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 population	0.391 (0.140)	0.316 (0.149)	0.344 (0.162)	1.043 (0.469)
Share of population 15–34	−0.995 (1.271)	−0.529 (1.544)	−3.504 (2.805)	−4.852 (4.678)
Observations	86	88	88	88
$R^2$	0.226	0.222	0.178	0.193
Number of clusters	43	44	44	44
Panel D: Panel regressions controlling for age structure, 1940s–1980s				
log population	0.331 (0.124)	0.313 (0.153)	0.265 (0.148)	0.968 (0.442)
Share of population 15–34	−1.068 (0.490)	−1.050 (0.599)	−2.095 (1.432)	−4.102 (1.858)
Observations	227	228	228	228
$R^2$	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 (4.4) 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. In computing standard errors, Bangladesh, India, and Pakistan are considered a single cluster. See the text and Supplementary Appendix Table A-1 for definitions and details.

0.3 more years of conflict in the 1980s (relative to the 1940s). Alternatively, this estimate implies that the average change in log population in our sample of 0.676 will be associated with 2.18 more years of conflict.<sup>11</sup>

11. 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). Similarly, the 0.3 number in the previous sentence is obtained as  $0.323 \times 0.10 \times 10 \approx 0.3$ .

The magnitude of this coefficient is fairly stable across different conflict datasets, as seen in columns 2 and 3, which use the UCDP/PRIO and Fearon-Laitin datasets respectively. To check the robustness of our results across alternative measures of conflict, column 4 considers  $\log(1 + \text{battle deaths per initial population})$  as the dependent variable. The coefficient for population remains positive and significant at 10%, and implies similar quantitative effects. Panel B shows analogous results from estimating equation (4.4) using panel data.

One possible concern with the results in Panels A and B is that they might be driven by age composition effects. In particular, rather than *larger* populations being associated with more conflict, it may be that *younger* populations are an important causal factor (Urdal, 2006). Panels C and D address this issue by including the share of population from 15 to 34 years of age as an additional independent variable (we lose eight countries due to lack of data). Although the share of young people, which is endogenous to population growth, is a “bad control” (a term from Angrist and Pischke, 2008), this specification is nonetheless useful for verifying whether there is a correlation between change in population and conflict beyond what is accounted for by the presence of larger young cohorts. The results are consistent with Panels A and B—the coefficients and significance level for log population are similar. The point estimate for the share of young population is negative and sometimes marginally 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.<sup>12</sup>

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. For example, if population increases coincide with general improvements in economic conditions or the institutional environment, their OLS relationship with conflict will be biased downward. We next investigate this issue with our instrumental variables strategy.

#### 4.3. *International epidemiological transition*

Our identification strategy relies on the *international epidemiological transition*, which generated significant improvements in health conditions and life expectancy across a wide range of countries. At the root of this transition were: major innovations in drugs (*e.g.* penicillin, other antibiotics) and associated effective treatments; chemicals (*e.g.* DDT); and broader public health campaigns spearheaded by international agencies such as the World Health Organization, which spread best practices around the world. These interventions had greater impact on countries where infectious diseases were initially more prevalent. Because this wave of innovations and campaigns was not caused by conditions in the countries of our sample, this episode provides a potentially exogenous source of variation in health conditions and thus in population in both our cross-country dataset and within Mexico.

For the cross-country analysis, we use the *predicted mortality instrument* from Acemoglu and Johnson (2007). This measures is the sum of each country’s initial (1940) mortality rate from infectious diseases until the moment of a global intervention (an innovation or campaign). After a global intervention occurs, the mortality rate from the disease in question declines to the lowest or “frontier” mortality rate in our sample. For country  $i$  at time  $t$ , the instrument is defined as

$$M_{it}^I = \sum_{d \in D} ((1 - I_{dt})M_{di40} + I_{dt}M_{dFt}), \quad (4.5)$$

12. In Supplementary Appendix Table A-3, we use the share of the population aged 20–39 as an alternative and obtain very similar results.

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  $t$  (taken to be zero mortality as in [Acemoglu and Johnson, 2007](#)); and  $D$  is the set of 13 diseases we use.<sup>13</sup>

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 had higher mortality from infectious disease in 1940 are expected to experience larger increases in population after the intervention.<sup>14</sup>

We use this variable as an instrument for population. Specifically, we posit the first-stage relationship for country  $i$  at time  $t$ ,

$$x_{it} = \varphi M_{it}^I + \tilde{\zeta}_i + \tilde{\mu}_t + \mathbf{Z}'_{it} \tilde{\beta} + u_{it}, \quad (4.6)$$

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

[Acemoglu and Johnson \(2007\)](#) show that changes in predicted mortality lead to major improvements in life expectancy and also significantly increased births as more women survived to childbearing age. In countries such as India, Pakistan, Indonesia, Ecuador, and El Salvador, where predicted mortality declined by a large amount, there were big gains in life expectancy and consequently sizable increases in population. In contrast, in parts of Western Europe, Uruguay, Argentina, South Korea, and Australia, where predicted mortality did not decrease as much, life expectancy remained comparatively unchanged and there were smaller increases in population.

The exclusion restriction for our IV strategy,  $\text{Cov}(M_{it}^I, \varepsilon_{it}) = 0$ , where  $\varepsilon_{it}$  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. We provide a series of falsification exercises suggesting that predicted mortality is not correlated with various pre-existing determinants of economic growth and conflict. Another important concern is that changes in predicted mortality may affect conflict not just through population but via changes in overall health conditions, age composition, or incentives to acquire education. We provide some evidence against these specific channels below, but this concern needs to be borne in mind in interpreting our IV estimates.

#### 4.4. First stages, reduced forms, and falsification

Table 3 (and Figure 1) takes a first look at the first-stage relationship between population and predicted mortality, as well as reduced-form regressions and falsification tests. We run the following long-differences regression (equivalent to a panel regression with two data points):

$$\Delta y_{it_1, t_0} = \alpha + \varphi \Delta M_{i1980, 1940}^I + \varepsilon_{it}, \quad (4.7)$$

13. These 13 diseases are (roughly in descending order of importance): malaria, pneumonia, tuberculosis, influenza, cholera, typhoid, smallpox, whooping cough, measles (rubeola), diphtheria, scarlet fever, plague and typhus. [Acemoglu and Johnson \(2007\)](#) discuss 15 diseases, but we do not have sufficient coverage for dysentery, and yellow fever does not add any useful variation in our sample.

14. The predicted mortality instrument depends on the choice for dating global interventions. An alternative “global mortality instrument” employed in [Acemoglu and Johnson \(2007\)](#) is independent of the coding of global interventions, assuming instead that each country’s initial mortality rate decreases at the pace of the global mortality rate for the disease in question. Our results are robust to using such alternative, as shown in Table A-2.

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

In columns 1 and 2, the dependent variable is the change in log population from 1940 to 1980. Column 1 includes all countries in the base sample, and column 2 focuses on low- and middle-income countries. In both cases, we observe a strong negative first-stage relationship between log population and predicted mortality (illustrated in Figure 1(a) and (b)).<sup>15</sup> The coefficient estimate of  $\varphi$  in column 1,  $-0.782$ , is significant at less than 1% and sizable. It 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—approximately a 45% increase in total population. The mean population in our sample in 1940 was about 34.7 million, so this estimate translates into an increase in population of 12.8 million. 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 half of the increase in population between 1940 and 1980.

In columns 3 and 4 (and Figure 1(c) and (d)), we turn to the relationship between predicted mortality and the fraction of each decade in conflict from 1940 to 1980. For the base sample and for low- and middle-income countries, we see 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 in columns 1 and 2, this translates into a positive effect of population on conflict as our 2SLS estimates below show.

If this causal interpretation—from the international epidemiological transition to population and via this channel to conflict—is correct, we should not see a correlation between changes in predicted mortality and pre-1940 changes in population or conflict. We explore this falsification exercise in columns 5–8 of Table 3 by estimating a variant of equation (4.7) where  $t_1 = 1940$  and  $t_0 = 1900$ . Reassuringly, we find no such pre-trends, and the coefficient estimates are insignificant and very small relative to our reduced form estimates. Predicted mortality explains changes in population and conflict after 1940, but not before 1940.

These results offer further confirmation that there were no preexisting trends in population or conflicts related to changes in predicted mortality. This bolsters our confidence that the predicted mortality instrument is not capturing some other economic or political trends.

The scatter plots in Figure 1, which show these first stages, reduced forms, and falsification exercises, further illustrate that the patterns in Table 3 are not driven by outliers. In each scatter plot, we highlight the data point for Mexico to place our within-country case in the broader context of our cross-country analysis. Mexico lies around the middle of the distribution for changes in predicted mortality and features relatively high population growth. Interestingly, it experienced no change in conflict when measured as years in conflict in the decade (in fact, it did not experience a year with civil war in either 1940 or 1980). This reflects the high fatalities thresholds and the limited focus of conflict measurement in international datasets, but it masks variation in social conflict in the form of citizen protests, some of them quite violent. This limitation of cross-country data is yet another motivation for our within-country analysis.

#### 4.5. 2SLS results

Table 4 presents our main results, which are the 2SLS estimates of the effect of population on conflict. Our second-stage regression is given by equation (4.4), where population is instrumented

15. In Supplementary Appendix Tables A-8, A-11, and A-13, we show the panel versions of these reduced forms, as well as their robustness to alternative samples and to the inclusion of differential trends.



TABLE 3  
First stages, reduced forms and falsification exercises

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	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
	First stages			Reduced forms		Falsification exercises		
Dependent variable is:	log population change 1940s–1980s			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	-0.782 (0.141)	-0.764 (0.191)	-0.491 (0.179)	-0.660 (0.236)	0.085 (0.055)	0.197 (0.126)	-0.189 (0.138)	-0.198 (0.196)
Observations	51	40	52	41	36	28	52	41
R <sup>2</sup>	0.306	0.238	0.166	0.200	0.012	0.039	0.033	0.029

Notes: OLS regressions (equation (4.7) in the text). Robust standard errors (clustered by country) are reported in parentheses. In computing standard errors, Bangladesh, India, and Pakistan are considered a single cluster. See the text and Supplementary Appendix Table A-1 for definitions and details.

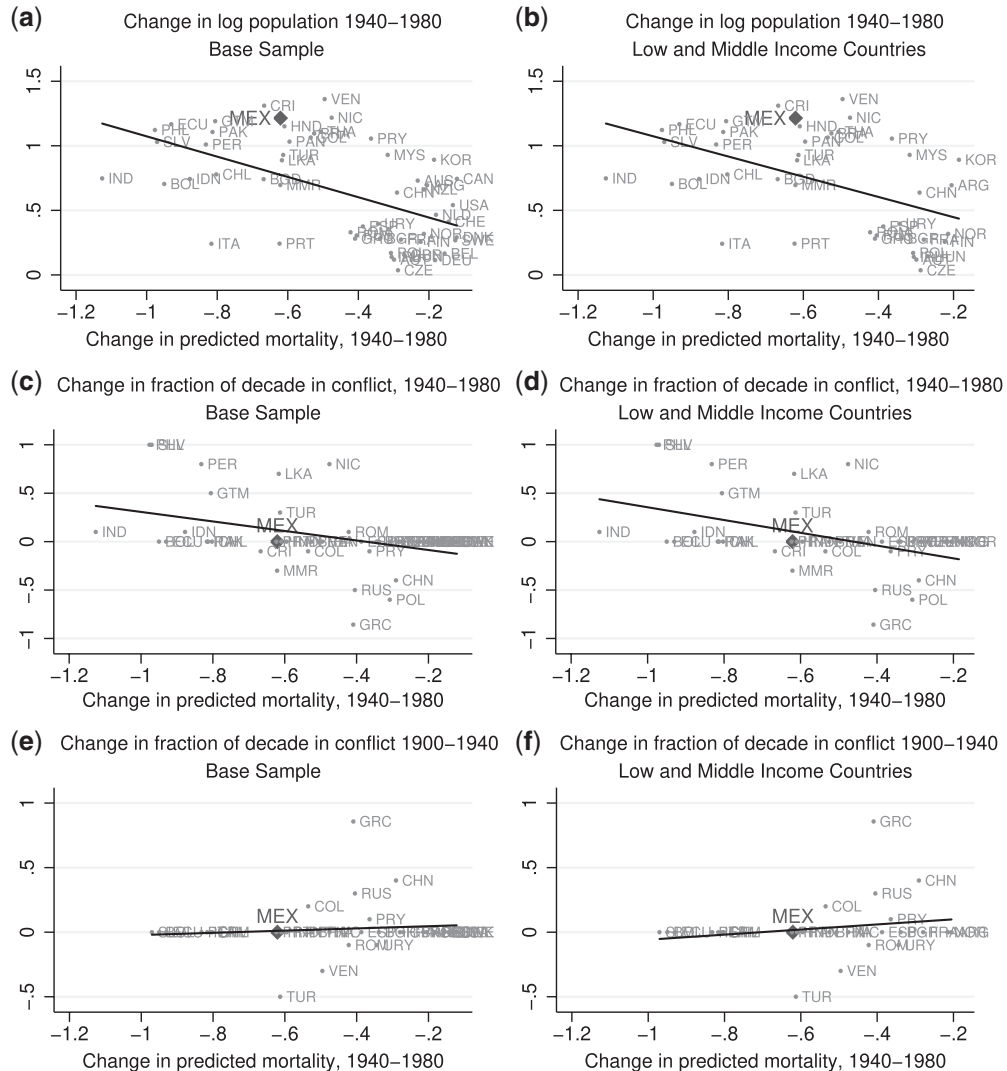


FIGURE 1

*Reduced forms, falsification, and first stages*

Notes: See Supplementary Appendix Table A-1 for variable definition and sources. Initially rich, initially middle, and initially poor are the top, middle, and bottom third set of countries in the base sample by level of income per capita.

by predicted mortality as in equation (4.6). We report long-difference regressions for 1940 and 1980 in Panel A and panel regressions for 1940–80 in Panel B. This table shows that the effect of population on conflict is positive and highly significant in almost all specifications.

In column 1, the dependent variable is the share of years in internal conflict per decade, as measured by the COW dataset. The coefficient of interest,  $\pi$ , is estimated at 0.617. This implies that a 10% increase in population leads to about 0.62 more years of conflict in the 1980s relative to the 1940s, or the average change in log population in our sample, 0.676, has been associated with a 4.2 more years in conflict during the 1980s (again relative to the 1940s).

We find similar results in our panel regressions for 1940–80 presented in Panel B ( $\pi = 0.609$ , significant at the 1% level). For example, for El Salvador, which experienced an increase in log

TABLE 4  
*The effect of population on conflict: 2SLS estimates*

	Dependent variable is fraction of decade in conflict (cols 1–3), and log(1 + battle deaths/pop. in 1940) in col 4			
	(1)	(2)	(3)	(4)
	COW	UCDP/PRIO	Fearon and Laitin	log(1+battle deaths/ pop. 1940)
Panel A: Long differences, just 1940s and 1980s				
log population	0.617 (0.213)	0.576 (0.238)	0.879 (0.303)	1.347 (0.598)
Observations	102	104	104	104
Number of clusters	50	51	51	51
Panel B: Panel regressions, 1940s–1980s				
log population	0.609 (0.205)	0.304 (0.250)	0.873 (0.461)	1.106 (0.454)
Observations	307	308	308	273
Number of clusters	63	63	63	54

*Notes:* 2SLS regressions with a full set of year and country fixed effects (equation (4.4) in the text, where population is instrumented by predicted mortality, as in equation (4.6) 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 stage, for the sample with data on years in conflict according to COW, in column 1 (panels B and D) of Table 5. In computing standard errors, Bangladesh, India, and Pakistan are considered a single cluster. See the text and Supplementary Appendix Table A-1 for definitions and details.

population of 0.46 during this period (a change in population from 1.6 to 4.6 million), the IV estimate of 0.617 implies an effect of roughly 2.8 more years in conflict ( $0.617 \times 0.46 \times 10$ ).

Both in long-differences and in panel specifications, the IV estimates are larger than the OLS results reported in Table 2. There are three reasons for OLS and IV estimates to differ in general. First, OLS estimates may be attenuated due to measurement error in the key right-hand side variable, in this case population. Though a possibility, we do not think this is a major concern in this case since population numbers are fairly accurate. Second, OLS and IV estimates capture different local average treatment effects. Third and most importantly, OLS estimates are affected by omitted variable biases, in particular, because changes in population are likely to be correlated with overall economic conditions and potential changes in the institutional environment, both of which could impact conflict. The signs of these correlations are *ex ante* ambiguous. The fact that OLS estimates are smaller than IV estimates suggests that these correlations are positive, leading to downward bias in OLS correlations.

Columns 2–4 investigate the robustness of the IV estimates. 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 the 1% level, with the exception of the UCDP/PRIO regressions in Panel B.<sup>16</sup>

16. We have a sparse distribution of countries around the globe, so strong spatial correlation between them is not a first-order concern, with the exception of India, Pakistan, and Bangladesh, which are treated as a single cluster. Moreover, our main results are robust to correcting the standard errors for spatial correlation. In Supplementary Appendix Table A-4, we report similar results when we apply Conley's (1999, 2010) correction for spatial correlation using a maximum radius of 9,684 km (the results are very similar with other distances).

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. We use battle deaths data to examine whether this may be driving our results. Column 4 considers log battle deaths for each reference date, divided by population in 1940, to calculate  $c_{it}$ . The coefficient on population is again positive and significant.<sup>17</sup>

#### 4.6. *Alternative samples, instruments, World War II and timing*

Table 5 presents estimates of our first- and second-stage relationships based on equations (4.4) and (4.6) for alternative samples and specifications. Panels A and B report long-difference specifications, and Panels C and D report panel regressions.

Column 1 includes all countries in our sample. In Panels A and C, this column reproduces our baseline estimates from column 1 of Table 4 to facilitate comparison. Columns 2 and 3 repeat the same set of regressions excluding Eastern and Western Europe, respectively. The estimates are broadly similar, suggesting that our results are not unduly affected by the various conflicts before and during World War II, which were concentrated in Europe. In Supplementary Appendix Table A-7, we also show that the results are very similar when we exclude Asia, Latin America, Australia, or Africa from our sample.

Column 4 drops initially rich countries to verify that the results are not driven by the comparison between rich and poor nations. Columns 5–7 check whether results are driven by events around World War II. Column 5 excludes the countries demographically most affected by that war, namely Austria, China, Finland, Germany, Italy, and the Russian Federation (Uralanis, 2003). Column 6 assigns the level of conflict of the 1950s to the 1940s. Column 7 drops the war years completely and assigns the number of years in conflict from 1946–9 (as a fraction of the 4 years in these interval) to our dependent variable in 1940. The results are similar in all cases.

Columns 8 and 9 investigate whether our IV results are driven by the impact of changes in predicted mortality on age composition or the overall health of the population. Column 8 controls for the share of the young (15–34) in the population. Despite being endogenous to changes in population, this variable is not statistically significant in our regressions and the effects of population on conflict are very similar to our baseline estimates. Column 9 controls for (log) life expectancy at 20 as a measure of the overall health of the adult population. The “bad control” concerns are more pronounced for this variable. Nevertheless, life expectancy at 20 is itself insignificant and the estimates of the impact of population on conflict are similar to our baseline results, even if less precise.<sup>18</sup>

Finally, column 10 extends the analysis to the 2000s. There is again a significant impact of population on conflict, though the size of the effect in the 2SLS estimate is about 50% smaller than

17. In columns 1 and 2 of Supplementary Appendix Table A-5, we show that the results are similar if we use the inverse hyperbolic sine transformation of the ratio of battle deaths to population rather than  $\log 1 +$  battle deaths. Columns 3 and 4 of the same table show that the results are also robust when we use a dummy for any conflict in the decade. Supplementary Appendix Table A-6, on the other hand, shows that there appears to be no effect of population growth on external wars or on other political outcomes, with the exception of state failure, which is itself likely to reflect the effects of civil wars and other severe internal conflicts. This pattern suggests that the main effect of rapid population growth may be on civil wars, civil conflict, and other internal tensions, but not on inter-state conflict and not even other aspects of political change in the country.

18. The “bad control” issue is somewhat less important for the share of the young population because, as we show in Supplementary Appendix Table A-9, predicted mortality does not have a major impact on the size of young cohorts (those between 15 and 34 or between 20 and 39).

TABLE 5  
The effect of population on conflict: basic robustness.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		Excluding Eastern Europe sample	Excluding Western Europe	Low and middle income countries only	Excluding most affected by WWII	Base sample assign 1950 to 1940	Base sample assign 1946–49 to 1940	Adding population 15–34 as covariate	Adding life expectancy as covariate	Just 1940s and 2000s
Panel A: 2SLS estimates. Dependent variable is fraction of decade in conflict according to Correlates of War (long differences, just 1940s and 1980s)										
log population	0.617 (0.213)	0.657 (0.240)	0.985 (0.363)	0.828 (0.294)	0.610 (0.215)	0.420 (0.198)	0.640 (0.226)	0.704 (0.228)	0.987 (0.550)	0.296 (0.144)
Panel B: First-stage estimates. Dependent variable is log population (long differences, just 1940s and 1980s)										
Baseline predicted mortality	-0.782 (0.141)	-0.700 (0.141)	-0.584 (0.168)	-0.764 (0.191)	-0.811 (0.140)	-0.782 (0.141)	-0.782 (0.141)	-0.848 (0.173)	-0.544 (0.233)	-1.174 (0.190)
R <sup>2</sup>	0.823	0.847	0.887	0.828	0.842	0.823	0.823	0.832	0.854	0.815
Observations (Panels A and B)	102	92	72	80	94	102	102	86	84	102
Number of clusters (Panels A and B)	50	45	35	39	46	50	50	43	41	50
Panel C: 2SLS estimates. Dependent variable is fraction of decade in conflict according to Correlates of War (panel regressions, 1940s–1980s)										
log population	0.609 (0.205)	0.649 (0.241)	1.036 (0.406)	0.862 (0.307)	0.615 (0.211)	0.276 (0.110)	0.649 (0.236)	0.526 (0.212)	1.020 (0.464)	0.467 (0.152)
Panel D: First-stage estimates. Dependent variable is log population (panel regressions, 1940s–1980s)										
Baseline predicted mortality	-0.464 (0.094)	-0.402 (0.093)	-0.336 (0.118)	-0.471 (0.131)	-0.476 (0.095)	-0.464 (0.094)	-0.464 (0.094)	-0.482 (0.118)	-0.315 (0.140)	-0.679 (0.121)
R <sup>2</sup>	0.792	0.819	0.885	0.814	0.815	0.792	0.792	0.757	0.810	0.774
Observations (Panels C and D)	307	278	228	252	279	307	307	227	202	437
Number of clusters (Panels C and D)	63	57	47	52	57	63	63	46	48	63

Notes: Panels A and C: 2SLS regressions with a full set of year and country fixed effects (equation (4.4) in the text, where population is instrumented by predicted mortality, as in equation (4.6) in the text). Panels B and D: OLS regressions with a full set of year and country fixed effects (equation (4.6) in the text). Robust standard errors (clustered by country) are reported in parentheses. Panels A and B present long-difference specifications with two observations per country, one for the initial date and one for the final date. Panels C and D present unbalanced panels with one observation per decade. In computing standard errors, Bangladesh, India and Pakistan are considered a single cluster. See the text and Supplementary Appendix Table A-1 for definitions and details.

our baseline estimate (or 25% smaller in our panel specifications).<sup>19</sup> In terms of the first-stage relationship, the estimates of  $\varphi$  in all of these cases are similar and significant at less than 1%.

Panels C and D have the same structure as in Panels A and B, now using a panel regression with decadal observations. The results are still highly significant and broadly similar. Supplementary Appendix Table A-8 verifies that our reduced-form regressions are robust to the same specification checks.

#### 4.7. Differential trends

We next explore the robustness of the first- and second-stage relationships presented so far to differential trends related to other factors. Specifically, we augment the first-stage equation with interactions between various baseline characteristics and year dummies, as indicated in the following equation:

$$x_{it} = \varphi M_{it}^I + \tilde{\zeta}_i + \tilde{\mu}_t + \sum_{t=1940}^{1980} \kappa_i' \bar{\omega}_t + u_{it}, \quad (4.8)$$

where  $\bar{\omega}_t = 1$  in year  $t$  and zero otherwise, and  $\kappa_i$  are “time-invariant” characteristics of country  $i$ . Similarly, we estimate the following second-stage equation:

$$c_{it} = \pi x_{it} + \zeta_i + \mu_t + \sum_{t=1940}^{1980} \kappa_i' \bar{\omega}_t + \varepsilon_{it}. \quad (4.9)$$

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); initial (1930) GDP per capita and population; initial (1930) civil war; and latitude and a malaria ecology index as geographic controls; share of young people (15–34) and share of adults (20–39); measures of natural resources;<sup>20</sup> and ethnic polarization,<sup>21</sup> which are often emphasized in the empirical literature on civil war.<sup>22</sup> Table 6 shows that our estimates of  $\varphi$  and  $\pi$  are similar

19. Supplementary Appendix Table A-10 examines how the response of conflict to population growth changed over time in more detail. Columns 1-5 in the two panels report long-differences and panel regressions where the initial time period is  $t = 1940$  and the final date is 1960, 1970, 1980, 1990, and 2000. The effects are weaker when we end the sample in 1960 or in 1970. This might be because the impact of population on conflict is delayed as it takes some time for the effects of population-induced scarcity to work out. The effects are also smaller when we extend the sample to the 1990s or 2000s. This is partly because of the end of the Cold War. Indeed, in column 6 in Panel B of Table A-10, we show that the effects of population on conflict are greater in countries where there was greater U.S. or Soviet involvement during the Cold War (as measured by the number of CIA and KGB interventions coded by [Berger et al., 2013](#)).

20. [Sachs and Warner \(1999\)](#) have popularized the share of natural resource exports in GDP. As [Ross \(2006\)](#) notes, however, 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 5 and 6, we focus instead on diamond production per capita (from [Humphreys, 2005](#)) and oil and gas rents per capita (from [Ross, 2006](#)). We found similar results with oil production per capita (also from [Humphreys, 2005](#)).

21. In the text, we focus on an index of ethnic polarization constructed by [Montalvo and Reynal-Querol \(2005\)](#) based on the theoretical work by [Esteban and Ray \(1994\)](#). We present results with measures of ethnic divisions in Supplementary Appendix Tables A-18–A-21.

22. We include a malaria ecology index as an additional control given the emphasis that it has received in some of the literature, which enables us to show that our results are not unduly affected by malaria (see also the results in the next subsection exploiting variation in one disease at a time). In our within-Mexico analysis, we use within country variation in malaria conditions as a key source of variation in mortality and population.



TABLE 6  
The effect of population on conflict: robustness to differential trends

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	
				Interaction of post-year dummies with...							
				Interaction of share population 15–34 and 20–39, 1940	Diamond production per capita in 1960	Oil and gas rents per capita in 1960	Ethnic polarization	Initial war in 1930s	Latitude	Malaria ecology index	
Panel A: 2SLS estimates. Dependent variable is fraction of decade in conflict according to Correlates of War (long differences, just 1940s and 1980s)											
log population	0.748 (0.257)	1.132 (0.427)	0.617 (0.215)	0.722 (0.236)	0.623 (0.213)	0.596 (0.204)	0.738 (0.294)	0.497 (0.224)	0.864 (0.344)	0.732 (0.258)	
p-value for post year dummy × variable indicated at the top of each column	0.106	0.0541	0.973	0.672	<0.01	0.0173	0.401	0.533	0.111	0.0106	
Panel B: First-stage estimates. Dependent variable is log population (long differences, just 1940s and 1980s)											
Baseline predicted mortality	-0.730 (0.167)	-0.600 (0.219)	-0.762 (0.126)	-0.836 (0.183)	-0.776 (0.138)	-0.808 (0.138)	-0.599 (0.148)	-0.935 (0.189)	-0.540 (0.143)	-0.704 (0.142)	
R <sup>2</sup>	0.826	0.833	0.844	0.800	0.842	0.841	0.868	0.805	0.860	0.833	
p-value for post year dummy × variable indicated at the top of each column	0.461	0.0952	<0.01	0.823	<0.01	<0.01	<0.01	0.931	<0.01	0.0701	
Observations (Panels A and B)	102	100	102	86	102	102	96	88	102	102	
Number of clusters (Panels A and B)	50	49	50	43	50	50	47	44	50	50	

(Continued.)

TABLE 6  
(Continued.)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
				Interaction of post-year dummies with...						
	Institutions	log GDP per capita in 1930	log population in 1930	Interaction of share population 15–34 and 20–39, 1940	Diamond production per capita in 1960	Oil and gas rents per capita in 1960	Ethnic polarization	Initial war in 1930s	Latitude	Malaria ecology index
	0.706 (0.252)	1.150 (0.457)	0.617 (0.211)	0.517 (0.206)	0.614 (0.204)	0.592 (0.197)	0.712 (0.303)	0.328 (0.135)	0.835 (0.349)	0.709 (0.233)
	0.165	0.0640	0.0160	0.647	<0.01	0.0958	0.562	0.0480	0.236	0.101
	Panel C: 2SLS estimates. Dependent variable is fraction of decade in conflict according to Correlates of War (panel regressions, 1940s–1980s)									
	log population									
	<i>p</i> -value for post year dummy × variable indicated at the top of each column									
	Panel D: First-stage estimates. Dependent variable is log population (panel regressions, 1940s–1980s)									
	Baseline predicted mortality									
	–0.437 (0.113)	–0.361 (0.146)	–0.446 (0.086)	–0.495 (0.120)	–0.460 (0.093)	–0.478 (0.092)	–0.337 (0.096)	–0.559 (0.126)	–0.306 (0.092)	–0.411 (0.094)
	0.810 <0.01	0.818 <0.01	0.798 0.0538	0.750 0.0701	0.801 <0.01	0.809 <0.01	0.849 <0.01	0.757 0.0529	0.844 <0.01	0.806 0.0376
	<i>R</i> <sup>2</sup>									
	<i>p</i> -value for post year dummy × variable indicated at the top of each column									
	307	267	265	223	300	300	277	244	300	307
	63	53	52	45	61	61	56	50	61	63

Notes: Panels A and C: 2SLS regressions with a full set of year and country fixed effects (equation (4.9) in the text, where population is instrumented by predicted mortality, as in equation (4.6) in the text). Panels B and D: OLS regressions with a full set of year and country fixed effects (equation (4.8) in the text). Robust standard errors (clustered by country) are reported in parentheses. Panels A and B present long-difference specifications with two observations per country, one for the initial date and one for the final date. Panels C and D present unbalanced panels with one observation per decade. In computing standard errors, Bangladesh, India, and Pakistan are considered a single cluster. See the text and Supplementary Appendix Table A-1 for definitions and details.

across specifications both in long-differences and the panel specifications. Only in one case, when we include time interactions with GDP per capita in 1930 in column 2, the second-stage estimate increases notably—for example, from 0.617 in Table 4 to 1.132 in the long-differences specification. The effects of additional baseline characteristics are explored in Supplementary Appendix Table A-12, again with similar results.

Overall, the results in Table 6 suggest that our estimates so far are unlikely to be confounded by differential trends related to baseline differences across countries.

#### 4.8. *Instrument robustness checks*

As recommended in recent work by Goldsmith-Pinkham *et al.* (2018), we investigated the relative importance of different diseases for our results. Specifically, Table A-14 reports the “Rotemberg weights” which decompose the Bartik estimator into a weighted sum of the just-identified instrumental variable estimators using each disease share as a separate instrument.

Pneumonia, tuberculosis, and malaria (and to a lesser extent, influenza, typhoid and measles) in the 1940s are the key drivers of our results in both the panel and long-difference specifications. This is in line with our prior based on mortality rates. Tables A-15 and A-16 report first stage and second-stage estimates with a number of different specifications and different disease compositions to verify that individual diseases are not responsible for our results.

The estimates are similar when we exclude the diseases with the highest Rotemberg weights, pneumonia, malaria, and tuberculosis. When we exclude all three at the same time in column 5, the first-stage coefficient changes significantly, but this is most likely due to power issues and the 2SLS estimate in Table A-16 is similar. In columns 7–9, we report estimates using just pneumonia, just malaria or just tuberculosis. In these cases, there are significant differences in the size of the 2SLS estimate (it is quite a bit smaller with pneumonia than with just tuberculosis), but in all specifications we continue to estimate a statistically significant impact of population on conflict. The overall findings are consistent regardless of which disease subsets we focus on.

#### 4.9. *Mean reversion*

In response to Acemoglu and Johnson’s (2007) finding of no effects of life expectancy on income, Bloom *et al.* (2014) argued that the level of life expectancy in 1940 affected subsequent growth rates and should be included on the right-hand side. Acemoglu and Johnson (2014) showed that controlling for initial life expectancy or mean reversion dynamics does not change the finding of no positive impact of life expectancy on income. Although we suspect these potential dynamics to be less relevant for conflict than for income, we explore their possible role in our results in Table 7. Our first strategy is to include the initial level of conflict or population interacted with time dummies in our decadal panel to flexibly control for mean reversion and the impact of past population levels. Our second strategy allows lagged conflict to linearly impact current conflict. Finally, we also present results from a nonlinear model that imposes the specification of convergence dynamics proposed by Bloom *et al.* (2014).

For comparison, column 1 reports our baseline panel estimates (as in column 1 in Table 4, Panel B). Column 2 restricts the sample to countries with available information on initial conflict, reducing the set of countries from 65 to 58. This has essentially no effect on our key point estimate, which changes from 0.609 to 0.606 and remains statistically significant at 1%.

Column 3 includes the interaction of initial war with a full set of year dummies. The coefficient on population changes only slightly, to 0.584 with a standard error of 0.181, even though the interactions between time dummies and initial conflict are jointly significant (the *p*-value, under 0.01, is reported at the bottom of the table). Columns 4 and 5 add lagged conflict on the right-hand

TABLE 7  
The effect of population on conflict: controlling flexibly for the impact of initial conflict and mean reversion, using panel data

	Dependent variable is fraction of decade in conflict according to Correlates of War								
	Baseline specification			Including initial war in 1940, interacted with time dummies			Including initial pop. in 1940, interacted with time dummies		
	2SLS (1)	2SLS (2)	2SLS (3)	2SLS (4)	GMM (5)	2SLS (6)	2SLS (7)	GMM (8)	GMM (9)
Initial war is fraction of decade in conflict in the 1940s									
log population	0.609 (0.205)	0.606 (0.207)	0.584 (0.181)	0.266 (0.107)	0.238 (0.105)	0.609 (0.207)	0.304 (0.143)	0.158 (0.084)	
Lagged conflict				-0.002 (0.082)	0.157 (0.114)		-0.049 (0.073)	0.180 (0.154)	
$\pi$									0.445 (0.038)
$\lambda$									-0.062 (0.008)
Observations	307	281	281	272	235	265	254	245	
Number of clusters	63	56	56	56		52			
Effect of an increase of 0.10 in log pop from 1940 to 1980	0.061	0.061	0.058	0.027	0.028	0.061	0.029	0.019	0.045
p-value for post year dummy $\times$ initial war			<0.01	<0.01	0.303	0.013	0.06	<0.01	
Moments					38			38	89
Hansen p-value					0.392			0.721	0.988
AR2 p-value					0.525			0.979	

Notes: Unbalanced panels with one observation per decade. Column 1 is our baseline 2SLS regressions with a full set of year and country fixed effects (equation (4.4) in the text, where population is instrumented by predicted mortality, as in equation (4.6) in the text). Column 2 restricts the sample to the set of countries for which there is available data on "initial" (in the 1940s) conflict, measured as the fraction of decade in conflict. Columns 3, 4, and 5 include a full set of year dummies interacted with initial war. Columns 6, 7, and 8 interact instead with initial population. Arellano and Bond's GMM estimators (columns 5 and 8) remove country fixed effects by taking first differences and then construct moment conditions using all predetermined lags of conflict and predicted mortality as instruments. It is estimated in two steps and is optimally weighted. Column 9 presents the weighted two-step GMM estimates of the Bloom et al. (2014) proposed equation (3):  $\Delta c_{it} = \alpha_t + \pi \Delta \log \text{population}_{it} + \lambda \pi \log \text{population}_{i,t-1} - \lambda c_{i,t-1} + \varepsilon_{it}$ . In column 9, the second and all longer lags of conflict and the first and all longer lags of predicted mortality are used as instruments. Robust standard errors corrected for arbitrary serial correlation clustered at the country level (Bangladesh, India, and Pakistan are considered a single cluster) are reported in columns 1-4 and 6-7, and robust standard errors are reported in columns 5, 8, and 9. See the text and Supplementary Appendix Table A-1 for definitions and details.

side, allowing for mean reversion in conflict. Column 4 uses the standard 2SLS estimator and column 5 presents Arellano and Bond's (1991) optimally weighted two-step generalized method of moments (GMM) estimator, with predicted mortality as the external instrument. This further reduces the sample because it requires additional predetermined lags of conflict for estimation. Lagged conflict is not significant in either column, suggesting that mean reversion is not a major issue. Although the point estimate for population falls (to 0.266 with a standard error of 0.107 in column 4 and 0.238 with a standard error of 0.105 in column 5), the results are again broadly consistent with a positive and significant effect of population on conflict.

Columns 6, 7, and 8 modify the specifications of columns 3, 4, and 5 to include interactions of initial population (rather than initial war) with a full set of time dummies. Our key point estimate in column 6 is very close to the baseline estimate of 0.609 and remains statistically significant at 1%. Controlling for lagged conflict has little impact on the standard 2SLS estimator (column 7) or on Arellano and Bond's GMM estimator (column 8), and lagged conflict is not significant in either column. We thus conclude that mean reversion appears unimportant and the effect of population remains positive and significant (although somewhat less precise in column 8).

Finally, for completeness in column 9 we consider the error correction model proposed by Bloom *et al.* (2014),

$$\Delta c_{it} = \alpha_t + \pi \Delta x_{it} + \lambda \pi x_{i,t-1} - \lambda c_{i,t-1} + \varepsilon_{it}.$$

In this error correction model, conflict adjusts slowly to its steady state value. To estimate this equation, we use the optimally weighted two-step GMM estimator with the second and all longer lags of conflict and the first and all longer lags of predicted mortality as instruments. We continue to estimate a positive and statistically significant (long-run) impact of log population on conflict, given by the parameter  $\pi$  (0.445 with a standard error of 0.038). Moreover, the point estimate for  $\lambda$  is negative ( $-0.062$  with a standard error of 0.008), and confirms that there is little mean reversion in conflict. At the bottom of the table, we present the long-run impact of a 10% increase in population on conflict from the various different models. Though the short-run effects vary across columns, long-run impacts are quite similar across the various specifications. In particular, in the error correction model estimated by GMM this long-run impact is 0.045 compared to 0.061 in our baseline model.

#### 4.10. Mechanisms

Our Malthusian model presented in Section 2 makes two additional testable predictions. First, population increases should lead to greater conflict related to natural resources (as opposed to non-resource related conflicts for which we should not see much of an effect). Second, it should be increases in population without corresponding improvements in productivity and thus GDP that lead to greater conflict. We now explore these two predictions. We follow Rustad and Binningsbø (2012) who have classified all internal armed conflicts in UCDP/PRIO as related or unrelated to natural resources. In Table 8, we use this classification to investigate the impact of population on resource and non-resource conflicts. Consistent with our expectations, column 1 shows a sizable impact of population on resource-related conflicts. Quantitatively, this estimate is larger than our baseline estimate in Table 4 (*e.g.* 0.918 for resource-related conflicts compared to our baseline estimate of 0.617 for all conflicts). In contrast, column 2 indicates that there is no relationship between population and conflict unrelated to natural resources—the estimate is small and insignificant.

Columns 3 and 4 provide additional evidence consistent with this interpretation by splitting the sample between countries with population density above or below the median in our sample. The effects of population on conflict are significantly larger among high density countries where we expect competition for resources to be more intense.

TABLE 8  
*The effect of population on conflict: the importance of natural-resource related conflicts, population density, economic growth and economic hardship: 2SLS estimates*

	Dependent variable is fraction of decade in...											
	Natural resource conflict		Non-resource conflict		High Pop. density		Low Pop. density		GDP growth		Negative price shocks	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)				
Panel A: Long differences, just 1940s and 1980s												
log population	0.918 (0.253)	0.094 (0.157)	0.930 (0.335)	0.330 (0.233)	0.620 (0.211)	0.016 (0.025)	0.453 (0.186)	0.164 (0.167)				
Observations	102	102	50	50	51	51	102	102				
Number of clusters	50	50	24	25	50	50	50	50				
Panel B: Panel regressions, 1940s–1980s												
log population	0.690 (0.199)	0.000 (0.226)	1.019 (0.304)	0.162 (0.183)	0.456 (0.158)	0.142 (0.090)	0.581 (0.167)	0.026 (0.137)				
Observations	308	308	134	133	306	306	307	314				
Number of clusters	63	63	26	27	63	63	63	63				

*Notes:* 2SLS regressions with a full set of year and country fixed effects (equation (4.4) in the text, where population is instrumented by predicted mortality, as in equation (4.6) in the text). Robust standard errors corrected for arbitrary serial correlation clustered at the country level (Bangladesh, India, and Pakistan are considered a single cluster). Natural-resource and Non-resource conflicts as classified for the UCDP/PRIO conflict dataset by Siri and Binningsbø (2012). In Panel A,  $\text{Conflict}_{\text{slow}}$  is equal to zero if the change in log GDP per capita exceeds the median in the sample (fast growers) and equal to the change in conflict from 1940 to 1980 (change in the fraction of the decade in conflict according to Correlates of War) if the change in log GDP per capita is smaller than the median in the sample (slow growers).  $\text{Conflict}_{\text{fast}}$  is defined analogously as equal to the change in conflict for fast growers and zero for slow growers. These are regressed on changes in log population from 1940 to 1980. In Panel B,  $\text{Conflict}_{\text{slow}}$  is equal to zero in each decade if the change in log GDP per capita in the decade exceeds the median in the sample (fast growers) and equal to the fraction of decade in conflict according to Correlates of War otherwise (slow growers).  $\text{Conflict}_{\text{fast}}$  is defined analogously as equal to conflict for fast growers and zero for slow growers.  $\text{Conflict}_{\text{high hardship}}$  and  $\text{Conflict}_{\text{low hardship}}$  are constructed similar to  $\text{Conflict}_{\text{slow}}$  and  $\text{Conflict}_{\text{fast}}$ , but based on whether the number of years with economic hardship is above or below the median of its distribution during the relevant sample period. We identify times of economic hardship as years in which an index of commodity price shocks (computed as the sum of commodity price changes weighted by country-specific export to population ratios) is negative. See the text and Supplementary Appendix Table A-1 for definitions and details.



Columns 5 and 6 turn to the second prediction from our theory. We code two new dependent variables. The first one, used in column 3,  $\text{Conflict}_{\text{slow}}$ , is equal to zero if the change in log GDP per capita from 1940 to 1980 is above the median in the sample and equal to the change in conflict from 1940 to 1980 otherwise. The second one,  $\text{Conflict}_{\text{fast}}$  which is used in column 4, is defined analogously and is equal to zero when growth is below the median and equal to the change in conflict from 1940 to 1980 otherwise. These two variables capture in a simple way the incidence of conflict in countries experiencing slow and fast growth. Consistent with our theoretical predictions, there is a strong effect on  $\text{Conflict}_{\text{slow}}$ , capturing the fact that among the slow-growing countries high population is associated with a significantly higher incidence of conflict. In contrast, there is no effect on  $\text{Conflict}_{\text{fast}}$ , which suggests that growing economies can absorb higher population without creating scarcity and thus higher conflict.

Growth during our sample period is potentially endogenous to changes in population, however. Columns 7 and 8 deal with this problem by focusing on variation coming from the world prices of commodities that a country exports. We use export data from the UN Comtrade dataset and data on world prices of 24 commodities since 1940 (from Grilli and Yang, 1988, and Pfaffenzeller *et al.*, 2007). Using these data we construct an index of commodity price shocks for each country and year as  $P_{it} = \sum_{j=1}^J \Delta p_{jt} \times \omega_{ji}$ , where  $\Delta p_{jt}$  is the yearly change in the price of commodity  $j$  and  $\omega_{ji}$  is the export per capita of commodity  $j$  in country  $i$  in the 1960s. We identify times of economic hardship for country  $i$  as years in which  $P_{it}$  is negative. We then construct the analogues of  $\text{Conflict}_{\text{slow}}$  and  $\text{Conflict}_{\text{fast}}$  variables in columns 5 and 6,  $\text{Conflict}_{\text{high hardship}}$  and  $\text{Conflict}_{\text{low hardship}}$ , based on whether the number of years with hardship is above or below the median of its distribution during the relevant sample period. Column 7 shows a large and precise impact of population on conflict when there are negative commodity price shocks ( $\text{Conflict}_{\text{high hardship}}$ ), while in column 8 there is no such impact on  $\text{Conflict}_{\text{low hardship}}$ . These results confirm that the effects of population on conflict are much more pronounced in the presence of economic hardship.

As noted in the Introduction, changes in mortality may also induce greater investments in education (Hansen and Strulik, 2017). These changes might in turn increase violence because of unmet aspirations (Krueger, 2004; Campante and Chor, 2012). To investigate this issue, in column 1 of Supplementary Appendix Table A-17 we examine whether educational attainment responds to population growth in our 2SLS specifications and find no evidence of a major impact of population on education. The remaining columns add educational attainment as an additional determinant of conflict (using different measures of conflict). Clearly, education is potentially a “bad control”, caused partly by changes in population. The fact that it is not predicted by instrumented log population partially alleviates this concern, but the results should still be interpreted with caution. In any case, we do not find any consistent or significant effect of education on conflict (either in long differences or in the panel specification), and the impact of population continues to be positive and significant (even if its magnitude is somewhat smaller in some specifications). Overall, we do not find any evidence that education is an important channel linking changes in mortality and conflict.

Finally, we explored the heterogeneity of the effects of population with respect to several other economic, political and social variables in Supplementary Appendix Tables A-18 and A-19 by interacting population with various baseline characteristics (including those used in Table 6) and in Tables A-20 and A-21 by splitting the sample into countries above or below the median value of each characteristic.<sup>23</sup> In summary, Supplementary Appendix Tables A-18 and A-19 present

23. These characteristics are: GDP per capita, ethnic polarization, ethnic fragmentation, religious polarization, religious fragmentation, average ethno-linguistic fragmentation, ethnic dominance, ethnic inequality, oil inequality,

some evidence that countries with greater ethnolinguistic fragmentation respond more strongly to population pressures. The interactions with the other characteristics are not significant or robust. In no cases do these interactions alter the main effects of population. When we split the sample, there are some differences in the magnitudes of the population effects, but these differences are neither statistically significant nor robust.

## 5. WITHIN-COUNTRY EVIDENCE: MEXICO

We next turn to an investigation of the within-country variation in Mexico. Our empirical strategy mimics the one we used in the cross country, except that instead of the 13 diseases we used there, we focus on variation from malaria in the baseline period and then trace the implications of major anti-malarial campaigns of the 1940s and 1950s.

### 5.1. *Malaria in Mexico*

Malaria was a major cause of death in Mexico in the first half of the 20th century, and in 1936 the fight against malaria was declared a matter of national interest. In line with the broader international epidemiological transition, the 1940s and 1950s marked an intensification of these efforts.<sup>24</sup>

The introduction of DDT dramatically increased the effectiveness and reach of these campaigns (Blancarte-Meléndez and de Jesús Cabrera-Palma, 1959). DDT was initially applied in parts of the country where malaria was endemic. But by 1948 this approach had been extended nationwide (Rabell *et al.*, 1986), and regular use of DDT was “fully generalized” by the second half of the 1950s (Díaz-Barriga *et al.*, 2003; Centro Nacional de Salud Ambiental, 2000).

Consequently, there were major impacts on morbidity and mortality within less than a decade (Bleakley, 2010; Pan-American Health Organization, n.a.). Malaria cases decreased from 41,000 in 1955 to 4,000 in 1960 (Díaz-Barriga *et al.*, 2003). Malaria accounted for about 8% of all deaths in Mexico in 1931. This fell to 5.2% in 1940. By 1960, the rate had declined dramatically, to 1.75% of total deaths. By 1970 malaria accounted for a negligible 0.01% of total deaths in Mexico (INEGI, 2015).

These averages mask large geographical variation. Areas where malaria was endemic benefited significantly more than others. Available state-level data for 1940 confirms this. Malaria deaths were particularly high in the states of Tabasco (24.1% of all deaths), Oaxaca (18.2%), Chiapas (18%), Campeche (14.7%), Guerrero (12.7%), Morelos (11.1%), and Veracruz (11%). In contrast, in Aguascalientes, Zacatecas, Guanajuato, Durango, México, Coahuila, Baja California Sur, Tlaxcala, Chihuahua, and the Distrito Federal deaths from malaria accounted for no more than 1% of total deaths (INEGI, 1941). By 1960, malaria deaths (as a fraction of all deaths) had come down sharply in the high-malaria states and stood at: 7.3% in Tabasco, 8.9% in Oaxaca, 4.8% in Chiapas, 0.8% in Campeche, 3.5% in Guerrero, 0.6% in Morelos, and 4.7% in Veracruz.

country area, agricultural climatic suitability, agricultural soil suitability, average combined agricultural suitability (the product of the soil and climate components), crop suitability index, agriculture as a percentage of GDP (value added), educational attainment, and initial institutions (constraints on the executive and democracy scores).

24. Díaz-Barriga *et al.* (2003) note that anti-malaria campaigns in Mexico followed the general expansion of DDT (Dichlorodiphenyltrichloroethane) usage as an insecticide, following the successful application of DDT by the U.S. military in the Pacific during WWII. More context on international campaigns is provided by Alilio *et al.* (2004) and Pampana (1954).

### 5.2. Descriptive statistics

Descriptive statistics for our Mexico sample are provided in Table 9, both for the entire sample of Mexican municipalities and separately for those with malaria suitability (or equivalently predicted mortality) above or below the median.

From 1960 to 1969 there were, on average, 2.3 violent protests per hundred thousand initial (1940) inhabitants, but crucially this number is larger in areas that have above-median malaria suitability (3.0 versus 1.64). There are no similar differences for non-violent protests (12.04 in high malaria areas and 11.94 in low malaria areas).

The gap between high and low malaria suitability municipalities is most pronounced for protests related to natural resources (1.52 versus 0.54), while there is a smaller contrast for other protests (1.47 versus 1.1). There is no clear gap in historical conflicts (0.041 versus 0.055).

Finally, the table shows that population levels in 1940 are similar between areas with high and low malaria suitability, but subsequently, there is faster population growth in areas with high suitability to malaria: the average increase in log population from 1940 to 1960 is 0.46 in municipalities with above median suitability and 0.36 in municipalities below the median. This is in line with our expectation that anti-malarial campaigns increased life expectancy and births in parts of Mexico previously suffering from high malaria incidence.

### 5.3. Empirical strategy

We use a similar identification strategy as in our cross-country analysis, but focusing on the role of malaria. This is mainly because we do not have data on mortality from different diseases at the municipality level in Mexico, but recall that malaria was a major source of differential changes in mortality and population during this time period. We construct a predicted mortality instrument,  $\tilde{M}_{i,t}$ , across municipalities in Mexico, where in the base period (1940)  $\tilde{M}_{i,t}$  is equal to malaria suitability. With the same logic as our cross-country predicted mortality instrument, after the anti-malarial campaigns are underway, in particular in 1960, we set  $\tilde{M}_{i,t} = 0$  in all municipalities. Therefore, our within-Mexico predicted mortality instrument uses only information from baseline distribution of malaria suitability.

For our dependent variable,  $c_{i,t}$ , we exploit the fact that in the 1940s there were essentially no protests against the PRI, and we have reliable data on protests in the 1960s (where  $c_{i,t}$  is normalized by population as in our cross-country work). Thus, our 2SLS model mimics the cross-country models, in particular the second-stage equation is (4.4) instrumented by (4.6), that is,

$$c_{i,t} = \pi x_{i,t} + \zeta_i + \mu_t + \mathbf{Z}'_{i,t} \beta + \varepsilon_{i,t}, \quad (5.10)$$

$$x_{i,t} = \varphi \tilde{M}_{i,t} + \tilde{\zeta}_i + \tilde{\mu}_t + \mathbf{Z}'_{i,t} \tilde{\beta} + u_{i,t}, \quad (5.11)$$

where the only difference from our cross-country equations is that the predicted mortality variable,  $\tilde{M}_{i,t}$ , is based on a single disease, malaria, and uses the index of malaria suitability (rather than baseline mortality rates). In addition,  $x_{i,t}$  is log population in year  $t$ , now for municipality  $i$ , and municipality fixed effects and time fixed effects are defined as in our cross-country equations. The vector  $\mathbf{Z}_{i,t}$  again designates the covariates and now includes interactions between dummies for Mexican states and the post anti-malarial campaign dummy. These interactions partially control for the substantial cross-state differences within Mexico. Since spatial spillovers and correlation are a greater concern for smaller geographic units, in our within-Mexico work we always allow for spatial correlation between municipalities (Conley, 1999, 2010).<sup>25</sup>

25. We use a radius of 35.9 km for the distance cut-off. This corresponds to the 80th percentile of the distribution of the average distance between a municipality and its closest neighbour, and thus ensures that the error terms for most municipalities are allowed to be correlated at least with their first-order neighbours.

TABLE 9  
Mexico: descriptive statistics

Variable	Year	By declines in predicted mortality (Mexico)			Variable	Year	By declines in predicted mortality (Mexico)			
		Base sample (2)	Above median (3)	Below median (4)			Base sample (6)	Above median (7)	Below median (8)	
		<b>Social conflict</b>					<b>Disease</b>			
Violent protests	1960	2.314 (11.801)	2.995 (13.690)	1.637 (9.517)	Baseline predicted mortality (Mexico)	1940	0.426 (0.282)	0.668 (0.160)	0.184 (0.125)	
Violent natural resource protests	1960	1.024 (8.849)	1.516 (11.162)	0.535 (5.638)			<b>Droughts</b>			
Violent non- resource protests	1960	1.283 (7.066)	1.469 (7.055)	1.099 (7.074)	Non-harvest droughts	1960	7.179 (5.628)	7.114 (5.716)	7.243 (5.541)	
Non-violent protests	1960	11.987 (39.049)	12.039 (34.597)	11.935 (43.036)	Harvest droughts	1960	1.994 (3.136)	1.709 (2.654)	2.279 (3.531)	
Historical conflict	1860–1940	0.048 (0.241)	0.041 (0.206)	0.055 (0.272)			<b>Education</b>			
		<b>Population</b>			Literacy rate	1940	0.314 (0.192)	0.286 (0.190)	0.343 (0.190)	
log population	1920	7.991 (1.099)	7.897 (1.071)	8.083 (1.119)	Literacy rate	1960	0.509 (0.178)	0.469 (0.178)	0.550 (0.168)	
log population	1930	8.140 (1.075)	8.054 (1.023)	8.225 (1.118)	Literacy rate, 15–39	1940	0.371 (0.222)	0.334 (0.218)	0.410 (0.219)	
log population	1940	8.279 (1.095)	8.188 (1.049)	8.370 (1.132)	Literacy rate, 15–39	1960	0.592 (0.218)	0.542 (0.217)	0.642 (0.207)	
log population	1960	8.690 (1.191)	8.645 (1.145)	8.734 (1.234)	Primary school enrolment	1940	0.031 (0.049)	0.026 (0.043)	0.036 (0.055)	
Share of population 15–34	1940	0.318 (0.024)	0.324 (0.023)	0.311 (0.023)	Primary school enrolment	1960	0.079 (0.088)	0.066 (0.081)	0.093 (0.093)	
Share of population 15–34	1960	0.309 (0.028)	0.312 (0.028)	0.306 (0.027)						
Share of population 20–39	1940	0.289 (0.025)	0.295 (0.025)	0.283 (0.023)						
Share of population 20–39	1960	0.264 (0.033)	0.266 (0.037)	0.261 (0.027)						

Notes: Municipal-level information. The table reports the mean values of variables in the samples described in the column heading, with standard deviations in parentheses. Predicted mortality (Mexico) is equal to malaria suitability in 1940 and to zero in 1960. Protests are counts of news stories about protests, expressed as a fraction of baseline population (per 100,000 people). Historical conflicts are battles fought in each municipality from 1816 to 1940. Columns 3 and 4 report descriptive statistics for subsamples in which malaria suitability was above or below the median value in the base sample (0.39). See the text and Supplementary Appendix Table A-1 for more details and definitions.

#### 5.4. First stages, reduced forms and falsification

We start by estimating the first-stage relationship summarized in equation (5.11) in column 1 of Table 10.<sup>26</sup> Panel A of the table focuses on bivariate relationships without any controls (except state fixed effects). Column 1 shows that the within-Mexico predicted mortality variable, based

26. As in equation (4.7) in the previous section, we report these in the form of long differences, which are algebraically equivalent to equation (5.11) with two data points.

on malaria suitability, is a strong predictor of municipality population growth. The coefficient of  $\tilde{M}_{i,t}$  is  $-0.287$  with a standard error of  $0.042$ . This implies that moving from a municipality that had no suitability for malaria in the base period to one with average suitability ( $0.43$ ) increases log population by roughly  $0.12$  ( $0.43 \times -0.29$ ), or equivalently, it raises population by about  $12\%$ . Column 2 presents the reduced-form between violent protests and our predicted mortality measure, revealing a strong and significant relationship (the coefficient is  $-3.52$  with a standard error of  $1.24$ ). In this case, the move from no suitability to average suitability is associated with  $1.5$  additional protests. Figure 2 depicts the first-stage and reduced-form relationships.<sup>27</sup>

In Panel B, we verify the robustness of these estimates to additional controls, which include basic geographic controls, initial population, measures of initial education, estimates of historical conflict (except in column 9), and measures resource abundance and ethnic composition (these controls are described in detail in the next subsection in the context of our discussion of Table 11). Both the first stage and the reduced form remain strong and statistically significant, even if the magnitudes of the coefficients decline somewhat when we add these controls (to  $0.21$  for the first stage and to  $2.75$  for the reduced form).

Columns 3–7 of Table 10 conduct a number of falsification exercises. We first explore whether our instrument also predicts population changes *before* 1940, specifically from 1930–1940 (column 3) or from 1920 to 1940 (column 4).<sup>28</sup> A significant correlation in these specifications would contradict our underlying assumption that the impact on population after 1940 comes from the anti-malarial campaigns rather than from other differential trends affecting municipalities with different levels of initial malaria. Reassuringly, our predicted mortality instrument is not robustly correlated with population growth in the preceding 10- or 20-year period, and the point estimates are much smaller than in column 1. Without controls (Panel A), the coefficient on our within-Mexico predicted mortality variable is  $-0.043$  with a standard error of  $0.024$  in column 3 and  $-0.065$  with a standard error of  $0.046$  in column 4. When controls are included in Panel B, the coefficient is further reduced to  $-0.001$  in column 3 and to  $0.02$  in column 4.

We also show in columns 5 and 6 of Table 10 that the change in predicted mortality is not correlated with previous changes (1930–1940) in either the share of the young adult population (ages 20–39) or literacy. Finally, in column 7 we explore whether historical violence and social conflict measures are higher in municipalities with different patterns of predicted mortality. For this exercise, we rely on information on historical battles in Mexico (from 1616 to 1940) digitized and geolocated by Ramos-Toro (2018) based on work by Clodfelter (2002). Bolstering confidence in our empirical strategy, the change in predicted mortality is not correlated with previous incidence of historical violence.

27. Supplementary Appendix Figure A-1 shows the geographical variation in malaria suitability (left panel), population growth (middle panel) and violent protests (right panel). We observe wide variation in malaria suitability across the country. In line with Figure 2, this variation is correlated with areas that experienced more rapid population growth and more intense protest activity following the anti-malaria campaign.

Supplementary Appendix Table A-22 presents some back of the envelope calculations suggesting that the magnitude of the first-stage relationship is plausible. In particular, if we incorporate the positive effects on births from lives saved from malaria (documented in Acemoglu and Johnson, 2007), assume the same relationship between comorbidity from malaria and other diseases in Mexico as in our cross-country sample, and suppose that municipality-level malaria deaths declined linearly to the national average after the anti-malaria campaign, we estimate that a municipality with average suitability to malaria should have experienced a  $9.41\%$  increase in population between 1940 and 1960 relative to a municipality with no malaria suitability. This is a little smaller than but in the same ballpark as our estimated  $12.47\%$  increase in population.

28. Because of the turmoil following the Mexican Civil War, the 1920 census was conducted in late 1921 and for this reason may be less reliable. This motivates our greater emphasis on data from the 1930 census for the falsification exercises.

TABLE 10  
*Mexico: first stage, reduced form and falsification exercises*

Dependent variable is:	First stage		Reduced form		Falsification exercises		
	log population change 1940s–1960s (1)	Violent protests change 1940s–1960s (2)	log population change 1930s–1940s (3)	log Historical change 1920s–1940s (4)	Share of population change 1930s–1940s (5)	Literacy rate, change 1930s–1940s (6)	Historical conflict (7)
Panel A: No controls							
Δ Predicted mortality (Mexico)	–0.287 (0.042)	–3.518 (1.239)	–0.043 (0.024)	–0.065 (0.046)	–0.001 (0.005)	0.021 (0.017)	–0.016 (0.020)
Observations	2,379	2,381	2,370	2,353	2,016	2,015	2,401
Panel B: Controls							
Δ Predicted mortality (Mexico)	–0.213 (0.044)	–2.754 (1.128)	–0.001 (0.030)	0.023 (0.047)	0.005 (0.005)	0.013 (0.015)	–0.014 (0.029)
Observations	1,789	1,786	1,781	1,771	1,717	1,716	1,789

*Notes:* Municipal-level OLS regressions with a full set of state fixed effects. Robust standard errors allowing for spatial correlation between municipalities within a radius of 35.9 km. Predicted mortality (Mexico) is equal to malaria suitability in 1940 and to zero in 1960. Δ denotes changes from 1940 to 1960. Controls are distance to capital, distance to big cities, land quality index, log population in 1940, primary school in 1940, university enrolment in 1940, share of a municipality's area on a sedimentary basin, share of indigenous 1940, and historical conflict (except in column 9). See the text and Supplementary Appendix Table A-1 for definitions and details.



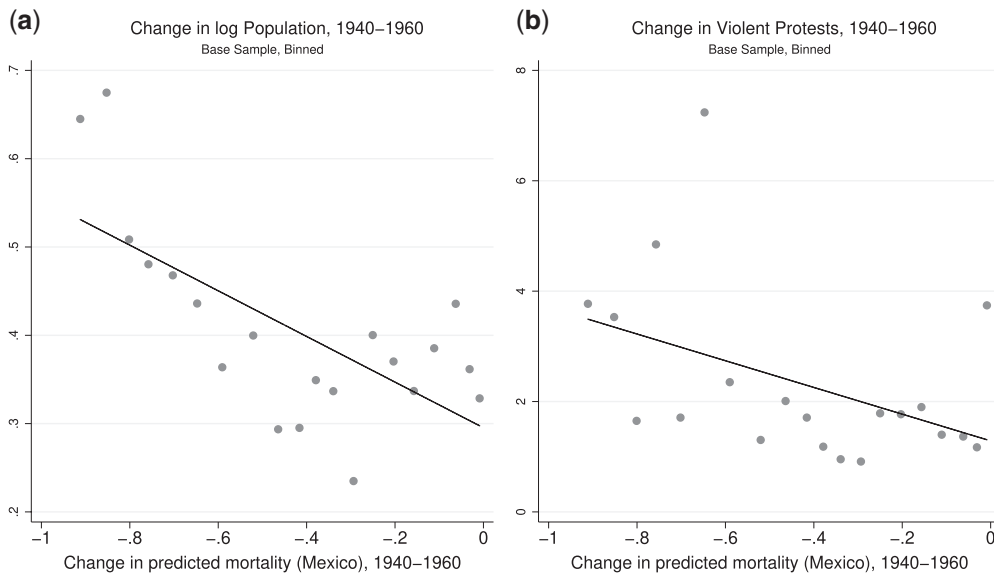


FIGURE 2

*Mexico: reduced form and first stage*

*Notes:* See Supplementary Appendix Table A-1 for variable definition and sources. Points represent averages within equal size bins constructed from the malaria suitability distribution. The fitted line shows the regression estimate.

### 5.5. Two-stage least squares results and robustness

In this subsection, we present our main 2SLS results for the impact of population on violent protests across Mexican municipalities between 1940 and 1960. Table 11 presents the 2SLS (Panel A), first-stage (Panel B), and OLS (Panel C) estimates.

Column 1 presents estimates of equation (5.10) without any additional interactions or controls. Both the OLS and IV estimates are positive and significant at less than 1%. Our more limited data for Mexico prevent us from controlling for the same set of baseline characteristics we explored in the cross-country analysis. Nevertheless, we have data on basic geographic features, initial population, measures of initial education, estimates of historical conflict, and proxies for resource abundance and ethnic diversity. In column 2 (and all subsequent columns) of Table 11, we include the following geographic controls: distance to Mexico City, distance to the nearest big city,<sup>29</sup> and a land quality index. In column 3, we also control for initial (1940) population. Column 4 controls for the initial level of education, using the primary school enrolment rate in 1940 (the results are similar when we use literacy rates for the entire population or for young cohorts). Column 5 uses the university enrolment rate as an alternative measure of education. Column 6 controls for the historical level of conflict. Column 7 controls for the share of a municipality's area on a sedimentary basin, which is associated with the presence of petroleum reserves and is unrelated to endogenous exploration efforts (see Cassidy, 2018). Column 8 controls for the share of people who only speak indigenous languages in the 1940 census. Finally, in column 9 we include all of these controls simultaneously.

29. This is defined as the linear distance from a municipality's centroid to the nearest municipality with a total population of at least 100,000 in 1960.

TABLE 11  
*Mexico: population and violent protests per capita. 2SLS, first stage and OLS estimates*

Dependent variable is violent protests (Panels A and C) and log population (Panel B)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: 2SLS estimates									
log population	12.372 (3.218)	13.333 (3.040)	12.813 (3.069)	13.102 (3.047)	12.389 (3.166)	12.996 (3.056)	14.263 (3.525)	10.948 (3.113)	13.049 (4.130)
Baseline control × post-anti-malaria campaign dummy									
log population, 1940			0.621 (0.220)						−0.108 (0.283)
Primary school, 1940				22.070 (6.101)					22.467 (7.709)
University enrolment, 1940					161.524 (90.083)				3.572 (43.801)
Historical conflict						3.021 (0.716)			1.746 (0.596)
Sedimentary basin							−0.617 (0.567)		−0.558 (0.830)
Share indigenous, 1940								−1.818 (0.752)	−1.029 (0.913)
Panel B: First-stage estimates									
Predicted mortality (Mexico)	−0.284 (0.030)	−0.262 (0.030)	−0.259 (0.030)	−0.247 (0.029)	−0.236 (0.030)	−0.259 (0.030)	−0.240 (0.032)	−0.246 (0.028)	−0.211 (0.031)
Panel C: OLS estimates									
log population	4.700 (0.841)	4.780 (0.853)	4.700 (0.843)	4.469 (0.880)	5.163 (0.914)	4.577 (0.850)	4.748 (0.858)	6.152 (1.033)	5.132 (1.067)
Observations	4,754	4,754	4,754	4,128	4,176	4,754	4,754	3,760	3,572
Geographic controls × post-anti-malaria campaign dummy		✓	✓	✓	✓	✓	✓	✓	✓

*Notes:* Municipality-level regressions with observations for 1940 and 1960. All regressions include a full set of municipality fixed effects as well as state fixed effects interacted with the post year dummy. Additional controls are indicated in Panel A (and also included in the other two panels but not reported to save space). Robust standard errors allowing for spatial correlation between municipalities within a radius of 35.9 km. Protests are counts of news stories about protests expressed as a fraction of baseline population (per 100,000 people). Predicted mortality (Mexico) is equal to malaria suitability in 1940 and to zero in 1960. Geographic controls are distance to capital, distance to big cities and land quality. See the text and Supplementary Appendix Table A-1 for definitions and details.

The coefficient on log population is statistically significant across all specifications, and its magnitude is relatively stable, ranging from 10.9 to 14.3 in the 2SLS specifications (Panel A). Moreover, the first stages reported in Panel B are robust across specifications, and the coefficient of the predicted mortality variable ranges from  $-0.28$  to  $-0.21$  and is always significant at less than 1%. Quantitatively, the 2SLS estimate in column 9 (13.05) implies that a 10% increase in population is associated with 1.3 more protests per hundred thousand inhabitants. Alternatively, the average change in log population in our sample of 0.41 is associated with 5.35 more protests. As in our cross-country estimates, the OLS coefficients in Panel C are smaller than the IV estimates. Our interpretation is again that the OLS estimates are biased downwards because of potential correlation between population changes and changes in other economic and institutional factors reducing conflict.

### 5.6. Mechanisms

We next investigate the same two implications from our conceptual framework which we explored in the context of cross-country data. These center on whether population increases lead to conflicts that are related to natural resources and whether these effects are exacerbated when there is economic hardship.

Columns 1–4 of Table 12 explore whether population increases are associated with resource-related protests. We use keywords to classify violent protests into those related to natural resources, such as water, land, mining, and agricultural products. Violent protests unrelated to any of these categories are coded as non-resource conflicts. Panel A of the table presents 2SLS estimates, while Panel B shows the corresponding first stages for log population (the first stages for the interactions in columns 7–10 are presented in Supplementary Appendix Table A-23). Column 1 focuses on resource-related (violent) protests and includes no controls other than state fixed effects, while column 2 adds our full set of controls as in column 9 of Table 11. The corresponding regressions for non-resource conflicts are in columns 3 and 4. The estimates show a robust effect of population on resource-related protests in columns 1 and 2. In contrast, the impact of population on non-resource protests is smaller and ceases to be statistically significant once our controls are added. The point estimates in columns 2 and 4, for example, imply that a 10% increase in population leads to a 0.87 more resource-related protests per 100,000 municipality inhabitants, while the same increase leads to only 0.43 more non-resource conflicts per 100,000 inhabitants.

Columns 5 and 6 turn to nonviolent protests as the dependent variable. There is a significant positive impact in column 5, but this is not robust to the same controls used in column 6. Overall, the evidence suggests that violent protests are more responsive to population pressures, though there appears to be some weaker effects on nonviolent protests as well.

Columns 7–10 investigate the second implication of our conceptual framework and investigate whether the effects of population pressures are greater when there are other sources of economic hardship. In particular, we exploit the random occurrences of drought across municipalities. Droughts are defined as precipitation levels below the 5th percentile of the long-run distribution (1900–2008) of monthly rain per municipality. As noted in Dell (2012), drought is harmful during most of the year because it lowers soil moisture content and reduces plant growth, but it is beneficial during the harvest season. We therefore focus on the number of months with droughts in the 1960s occurring during the non-harvesting period for corn. If a Malthusian mechanism links population and conflict as in our conceptual framework, then municipalities facing greater economic hardships due to droughts should experience greater increases in conflicts following population rises. Our estimates show that this is indeed the case and the interaction between drought and population is positive and significant, and is robust to the inclusion of our standard controls. We also include interactions for droughts in the harvest season as a placebo to investigate whether any other mechanisms aside from resources might be connecting droughts and violence.<sup>30</sup> These results are reassuring: the coefficient for this interaction is negative and not statistically significant.<sup>31</sup> The models in columns 9 and 10 additionally include interactions between the harvest and the non-harvest dummies and the post-anti-malarial comparing dummy as a control

30. Hsiang *et al.* (2013) argue that warmer temperatures and extreme rainfall influence conflict. Scheffran *et al.* (2012) discuss multiple possible causal pathways between weather and violence.

31. Droughts are demeaned when included in the interaction terms to facilitate interpretation of the direct effect of change in log population (which ensures that these effects are evaluated at the mean of the drought distribution). In Supplementary Appendix Table A-25, we verify the robustness of the results to alternative definitions of drought. For example, they are very similar when we define droughts using different percentiles of the long-run distribution of monthly rain. We also confirm that our results are robust to using the drought severity measure of Dell (2012), which is the maximum between 1 and the ratio of the decade's mean monthly precipitation to long-run average monthly precipitation.

TABLE 12  
*Mexico: population and protests per capita. 2SLS estimates heterogeneous effects and other outcomes*

Panel A: 2SLS estimates dependent variable is...	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Protests per capita by type						Violent protests per capita			
	Violent			Nonviolent			Interactions with non-harvest and harvest droughts			
	Natural resource	Natural resource	Non- resource	Non- resource	Nonviolent	Nonviolent				
log population	7.621 (2.667)	8.737 (2.839)	4.751 (1.630)	4.313 (2.686)	27.461 (8.785)	22.343 (13.928)	10.771 (3.623)	14.998 (5.045)	10.446 (3.924)	16.681 (4.987)
log pop × Non-harvest					0.446 (0.150)	1.173 (0.679)	0.446 (0.150)	0.542 (0.175)	1.173 (0.679)	1.681 (0.751)
log pop × Harvest					-0.295 (0.345)	-1.785 (1.318)	-0.295 (0.345)	-0.165 (0.458)	-1.785 (1.318)	-2.117 (1.519)
Panel B: First-stage estimates										
Dependent variable is log population										
Predicted mortality (Mexico)	-0.284 (0.030)	-0.211 (0.031)	-0.284 (0.030)	-0.211 (0.031)	-0.284 (0.030)	-0.211 (0.031)	-0.245 (0.037)	-0.204 (0.034)	-0.173 (0.046)	-0.157 (0.050)
Predicted mortality (Mexico) × Non-harvest							-0.007 (0.003)	-0.002 (0.003)	-0.016 (0.005)	-0.007 (0.005)
Predicted mortality (Mexico) × Harvest							0.014 (0.006)	0.007 (0.007)	0.016 (0.010)	0.007 (0.011)
Observations	4,759	3,572	4,759	3,572	4,756	3,572	4,744	3,566	4,744	3,566
Controls × post-anti-malaria campaign dummy		✓		✓		✓		✓		✓
Droughts × post-anti-malaria campaign dummy									✓	✓

*Notes:* Municipality-level regressions with observations for 1940 and 1960. All regressions include a full set of municipality fixed effects as well as state fixed effects interacted with the post year dummy. Population is instrumented with predicted mortality (Mexico). Robust standard errors allowing for spatial correlation between municipalities within a radius of 35.9 km. Protests are counts of news stories about protests in 1960, expressed as a fraction of baseline population (per 100,000 people). Predicted mortality (Mexico) is malaria suitability in 1940 and zero in 1960. A drought is defined as precipitation below the 5th percentile of the long-run distribution (1900–2008) of monthly rain per municipality and we count the number of months with droughts in the decade. Controls are distance to capital, distance to big cities, land quality index, log population in 1940, primary school in 1940, university enrolment in 1940, share of a municipality's area on a sedimentary basin, share of indigenous 1940 and historical conflict. Droughts are demeaned before interaction. See the text and Supplementary Appendix Table A-1 for definitions and details.

for any direct effects from droughts. This increases the standard errors of our drought interactions, but the overall pattern remains similar.

Overall, consistent with our conceptual framework and the results reported our cross-country analysis, it appears that increases in population caused a spike in violent protests across Mexican municipalities, and this impact was more pronounced in the presence of drought. Moreover, only droughts that affect yields (non-harvest droughts) raised the responsiveness of violent protests to population pressures.

As in our cross-country analysis, a final question is whether changes due to lower morbidity and mortality—resulting from the reduced incidence of malaria—are impacting conflict via changes in the share of younger cohorts in the population or through changes in education as opposed to our Malthusian mechanism.<sup>32</sup> To examine these issues, Panel A of Supplementary Appendix Table A-24 explores whether predicted mortality (Mexico) is correlated with changes in population shares or educational attainment. In Panel B, we use the share of the population aged 20–39 and educational attainment variables as controls in a regression for violent protests on (instrumented) log population. The population share and education measures are again likely “bad controls” because they are affected by overall changes in population, and in fact, in this case they do seem to be correlated with our instrument. All the same, our main effects remain largely unchanged when these variables are included.

## 6. CONCLUSIONS

The large population increases in many developing countries that followed the *international epidemiological transition* of the 1940s contributed to an increase in internal violent conflicts, including civil wars and violent protests. This pattern is present and highly robust both across countries and across Mexican municipalities.

Our conceptual framework proposes an explanation for these patterns based on a Malthusian channel—a larger population increases resource scarcity and encourages fighting over scarce resources. This conceptual approach highlights that the problem is not so much that higher population will always lead to greater conflict, but that population surges unaccompanied by corresponding increases in productivity or physical and human capital investments will do so. We bolster this interpretation by showing that our results are accounted for by increases in conflicts related to natural resources and the effects are significantly larger in economies experiencing other sources of economic hardship (slower growth or droughts).

Obvious questions for future research in this area include extending our analysis to other periods and samples, and investigating what economic changes and policy actions can better accommodate upcoming population surges around the developing world and avoid greater conflict.

*Acknowledgments.* We thank the editor Nicola Gennaioli and four anonymous referees as well as seminar participants at MIT’s development lunch, the University of Chicago, the XXII Annual Conference of the European Society for Population Economics at UCL, 2017 AEA Annual Meetings, Brown University, the Stanford-LSE-Uniandes 2019 conference on long-run development, Universidad de los Andes and Universidad del Rosario. Ioannis Tokatlidis and Juan Camilo Yamin provided superb research assistance. We also thank Alberto Díaz-Cayeros and Horacio Larreguy for comments and Jennifer Alix-García, Juan Felipe Riaño, Diego Ramos-Toro, and Emily Sellars for sharing their data.

### Supplementary Data

Supplementary data are available at *Review of Economic Studies* online.

32. We do not have data on life expectancy across municipalities to proxy for the general health of the population.

## REFERENCES

- ACEMOGLU, D., EGOROV, G. AND SONIN, K. (2012), "Dynamics and Stability of Constitutions, Coalitions and Clubs", *American Economic Review*, **102**, 1446–1476.
- ACEMOGLU, D. AND JOHNSON, S. (2007), "Disease and Development: The Effect of Life Expectancy on Economic Growth," *Journal of Political Economy*, **115**, 925–985.
- ACEMOGLU, D. AND JOHNSON, S. (2014), "Disease and Development: A Reply to Bloom, Canning and Fink," *Journal of Political Economy*, **122**, 1367–1375.
- ACEMOGLU, D. AND ROBINSON, J. A. (2001), "A Theory of Political Transitions," *American Economic Review*, **91**, 938–963.
- ACEMOGLU, D. AND ROBINSON, J. A. (2006), *Economic Origins of Dictatorship and Democracy* (New York, NY: Cambridge University Press).
- ALILIO, M. S., BYGBJERG, I. C. AND BREMAN, J. G. (2004), "Are Multilateral Malaria Research and Control Programs the Most Successful? Lessons from the Past 100 Years in Africa," *The American Journal of Tropical Medicine and Hygiene*, **71**, 268–278.
- ANGRIST, J. D. AND PISCHKE, J.S. (2008), *Mostly Harmless Econometrics: An Empiricist's Companion* (Princeton: Princeton University Press).
- ARELLANO, M. AND BOND, S.R. (1991), "Some Specification Tests for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, **58**, 277–298.
- ASHRAF, N., FINK, G. AND WEIL, D. N. (2010), "Evaluating the Effects of Large Scale Health Interventions in Developing Countries: The Zambian Malaria Initiative," (NBER Working Paper, 16069).
- BAROFSKY, J., ANEKWE, T. D. AND CHASE, C. (2015), "Malaria Eradication and Economic Outcomes in Sub-Saharan Africa: Evidence from Uganda," *Journal of Health Economics*, **44**, 118–136.
- BARLOW, R. (1967), "The Economic Effects of Malaria Eradication," *American Economic Review*, **57**, 130–148.
- BARTRA, A. (1985), *Los herederos de Zapata: Movimientos campesinos posrevolucionarios en México, 1920–1980* (México: Editorial ERA).
- BELLINGERI, M. (2003), *Del agrarismo armado a la guerra de los pobres, 1940–1974* (México: El Colegio de México, Centro de Estudios Internacionales).
- BERGER, D., EASTERLY, W., NUNN, N. AND SATYANATH, S. (2013), "Commercial Imperialism? Political Influence and Trade During the Cold War," *American Economic Review*, **103**, 863–896.
- BESLEY, T. J. AND PERSSON, T. (2008), "The Incidence of Civil War: Theory and Evidence," (NBER Working Paper, 14585).
- BEZDEK, R. (1973), "Electoral Oppositions in Mexico: Emergence, Suppression, and Impact on Political Processes," (PhD Dissertation, Ohio State University).
- BLANCARTE-MELÉNDEZ, J. R. AND DE JESÚS CABRERA-PALMA, N. (1959), "El programa de erradicación del paludismo en México," *Salud Pública de México*, **1**, 9–50.
- BLATTMAN, C. AND MIGUEL, E. (2010), "Civil War," *Journal of Economic Literature*, **48**, 3–57.
- BLEAKLEY, H. (2010), "Malaria Eradication in the Americas: A Retrospective Analysis of Childhood Exposure," *American Economic Journal: Applied Economics*, **2**, 1–45.
- BLOOM, D. E., CANNING, D. AND FINK, G. (2014), "Disease and Development Revisited," *Journal of Political Economy*, **122**, 1355–1366.
- BRÜCKNER, M. (2010), "Population Size and Civil Conflict Risk: Is There a Causal Link?," *Economic Journal*, **120**, 535–550.
- CAMERON, A. C., GELBACH, J. B. AND MILLER, D. L. (2008), "Bootstrap-based Improvements for Inference with Clustered Errors," *The Review of Economics and Statistics*, **90**, 414–427.
- CAMPANTE, F. R. AND CHOR, D. (2012), "Why was the Arab World Poised for Revolution? Schooling, Economic Opportunities and the Arab Spring," *Journal of Economic Perspectives*, **26**, 167–88.
- CASE, A. AND PAXSON, C. (2009), "Early Life Health and Cognitive Function in Old Age," *American Economic Review*, **99**, 104–09.
- CASSIDY, T. (2018), "The Long Run Effects of Oil Wealth on Development: Evidence from Petroleum Geology," (Mimeo, University of Alabama).
- CENTRO NACIONAL DE SALUD AMBIENTAL (2000), *Situación actual de la malaria y uso del DDT en México* (Centro de Vigilancia Epidemiológica, Secretaría de Salud).
- CLODFELTER, M. (2002), *Warfare and Armed Conflicts: A Statistical Reference, 1500–2000* (Jefferson: McFarland).
- COLELLA, F., LALIVE, R., SAKALLI, S.O., et al. (2018), *Inference with Arbitrary Clustering Mimeo* (University of Lausanne).
- COLLIER, P. AND HOFFLER, A. (1998), "On Economic Causes of Civil War," *Oxford Economic Papers*, **50**, 563–573.
- COLLIER, P. AND HOFFLER, A. (2004), "Greed and Grievance in Civil War," *Oxford Economic Papers*, **56**, 563–595.
- CONLEY, T. G. (1999), "GMM Estimation with Cross Sectional Dependence," *Journal of Econometrics*, **92**, 1–45.
- CONLEY, T. G. (2010), "Spatial Econometrics", in Durlauf, S. N. and Blume, L. E. (eds) *Microeconometrics* (London: Palgrave Macmillan), 303–313.
- DELL, M. (2012), *Path Dependence in Development: Evidence from the Mexican Revolution* (Cambridge, MA: Harvard University).
- DÍAZ-BARRIGA, F., BORJA-ABURTO, V., WALISZEWSKI, S. et al. (2003), "DDT in Mexico", in Fiedler, H. (ed.) *Persistent Organic Pollutants* (Berlin: Springer) 371–388.



- DIHEL, P. F. AND GLEDITSCH, N. P. (2001), *Environmental Conflict* (Boulder, CO: Westview Press).
- EASTERLY, W. AND LEVINE, R. (1997), "Africa's Growth Tragedy: Policies and Ethnic Divisions," *Quarterly Journal of Economics*, **112**, 1203–1250.
- ESTEBAN, J. AND RAY, D. (1994), "On the Measurement of Polarization," *Econometrica*, **62**, 819–851.
- ESTEBAN, J. AND RAY, D. (1999), "Conflict and Distribution," *Journal of Economic Theory*, **87**, 379–415.
- FAO/IIASA. (2010), *Global Agro-ecological Zones (GAEZ v3.0)* (Rome, Italy: FAO and Laxenburg, Austria: IIASA).
- FEARON, J. D. (1998), "Commitment Problems and the Spread of Ethnic Conflict", in Lake, G. and Rothchild, D. (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**, 275–301.
- FEARON, J. D. (2006), "Ethnic Mobilization and Ethnic Violence", In Weingast, B. R. and Wittman, D. A. (eds) *The Oxford Handbook of Political Economy* (Oxford and New York: Oxford University Press) 852–868.
- FEARON, J. D. AND LAITIN, D. D. (2003), "Ethnicity, Insurgency and Civil War," *The American Political Science Review*, **97**, 75–90.
- FERGUSON, L., LARREGUY, H. AND RIAÑO, J. F. (2018), "Political Competition and State Capacity: Evidence from a Land Allocation Program in Mexico" (Lacea Working Paper, No. 2018–0011).
- FLÜCKIGER, M. AND LUDWIG, M. (2017), "Malaria Suitability, Urbanization and Persistence: Evidence from China over More than 2000 Years," *European Economic Review*, **92**, 146–160.
- FUGRO ROBERTSON, Ltd. (2013), *Fugro Tellus Sedimentary Basins of the World Map* <http://www.datapages.com/gis-map-publishing-program/gis-open-files/global-framework/robertson-tellus-sedimentary-basins-of-the-world-map>.
- GARFINKEL, M. R. AND SKAPERDAS, S. (2007), "Economics of Conflict: An Overview", in Hartley, K. and Sandler, T. (Eds) *Handbook of Defense Economics* (Vol. 2) (Amsterdam, The Netherlands: Elsevier) 649–709.
- GETHING, P. W., VAN BOECKEL, T. P., SMITH, D. L., et al. (2011), "Modelling the Global Constraints of Temperature on Transmission of *Plasmodium falciparum* and *P. vivax*," *Parasites & Vectors*, **4**, 1–11.
- GLEDITSCH, N. P., WALLENSTEEN, P., ERIKSSON, M., et al. (2002), "Armed Conflict 1946–2001: A New Dataset," *Journal of Peace Research*, **39**, 615–637.
- GOLDSMITH-PINKHAM, P., SORKIN, I. AND SWIFT, H. (2018), "Bartik Instruments: What, When, Why, and How," (NBER Working Paper, w24408).
- GOLDSTONE, J. (1991), *Revolution and Rebellion in the Early Modern World* (Oakland, CA: University of California Press).
- GOLDSTONE, J. A. (2002), "Population and Security: How Demographic Change can Lead to Violent Conflict," *Journal of International Affairs*, **56**, 3–21.
- GOLDSTONE, J., KAUFMANN, E. AND TOFT, M. (2012), *Political Demography: How Population Changes are Reshaping International Security and National Politics* (Oxford: Oxford University Press).
- GRILLI, E. R. AND YANG, M. C. (1988), "Primary Commodity Prices, Manufactured Goods Prices, and the Terms of Trade of Developing Countries: What the Long Run Shows," *The World Bank Economic Review*, **2**, 1–47.
- HANSEN, C. W. AND STRULIK, H. (2017), "Life Expectancy and Education: Evidence from the Cardiovascular Revolution," *Journal of Economic Growth*, **22**, 421–450.
- HAUGE, W. AND ELLINGSEN, T. (1998), "Beyond Environmental Scarcity: Causal Pathways to Conflict," *Journal of Peace Research*, **35**, 299–317.
- HERRERA CALDERÓN, F. AND CEDILLO, A. (2012), *Challenging Authoritarianism in Mexico: Revolutionary Struggles and the Dirty War, 1964–1982* (New York: Routledge).
- HOMER-DIXON, T. F. (1991), "On the Threshold: Environmental Changes as Causes of Acute Conflict," *International Security*, **2**, 76–116.
- HOMER-DIXON, T. F. (1999), *Environment, Scarcity and Violence* (Princeton, NJ: Princeton University Press).
- HONG, S. C. (2007), "The Burden of Early Exposure to Malaria in the United States, 1850–1860: Malnutrition and Immune Disorders," *The Journal of Economic History*, **67**, 1001–1035.
- HSIANG, S. M., BURKE, M. AND MIGUEL, E. (2013), "Quantifying the Influence of Climate on Human Conflict," *Science*, **341**, 1212.
- HUMPHREYS, M. (2005), "Natural Resources, Conflict and Conflict Resolution," *Journal of Conflict Resolution*, **49**, 508–537.
- INEGI. (1941), *Anuario estadístico de los Estados Unidos Mexicanos 1941* (México: Instituto Nacional de Estadística y Geografía).
- INEGI. (2015), *Estadísticas Históricas de México 2014: 4. Salud* (México: Instituto Nacional de Estadística y Geografía).
- KAPLAN, R. (1994), *The Coming Anarchy* (Boston, MA: The Atlantic).
- KRUEGER, A. (2004), *What Makes a Terrorist: Economics and the Roots of Terrorism* (Princeton: Princeton University Press).
- LACINA, B. AND GLEDITSCH, N. P. (2005), "Monitoring Trends in Global Combat: A New Dataset of Battle Deaths," *European Journal of Population*, **21**, 145–166.
- LA PORTA, R., LOPEZ-DE SILANES, F., SHLEIFER, A., et al. (1999), "The Quality of Government," *Journal of Law, Economics and Organization*, **15**, 222–279.
- LUCAS, A. M. (2010), "Malaria Eradication and Educational Attainment: Evidence from Paraguay and Sri Lanka," *American Economic Journal: Applied Economics*, **2**, 46–71.
- MADDISON, A. (2003), *The World Economy: Historical Statistics* (Paris, France: Development Centre of the Organisation for Economic Co-operation and Development).

- MADDISON, A. (2006), *The World Economy* (Paris, France: Development Centre of the Organisation for Economic Co-operation and Development).
- MALTHUS, T. R. (1798), *An Essay on the Principle of Population, As IT Affects the Future Improvement of Society: With Remarks on the Speculations of Mr. Godwin, M. Condorcet and Other Writers* (London: Printed for J. John).
- MATSUURA, K. AND WILLMOTT, C. (2009), *University of Delaware Global Surface Air Temperature and Precipitation Climatology and Monthly Grids, 1900–2008*. <http://research.jisao.washington.edu/data/ud/>.
- MICHALOPOULOS, S. (2012), “The Origins of Ethnolinguistic Diversity,” *American Economic Review*, **102**, 1508–1539.
- MIGUEL, E., SATYANATH, S. AND SERGENTI, E. (2004), Economic Shocks and Civil Conflict: An Instrumental Variables Approach. *Journal of Political Economy*, **112**, 725–753.
- MONTALVO, J. G. AND REYNAL-QUEROL, M. (2005), “Ethnic Polarization, Potential Conflict and Civil Wars,” *American Economic Review*, **95**, 796–816.
- MUELLER, H. (2016), “Growth and Violence: Argument for a Per Capita Measure of Civil War,” *Economica*, **83**, 473–497.
- PAN-AMERICAN HEALTH ORGANIZATION. (n.a.), *Antología de la Atención a la Salud en México, 1902–2002*. [https://www.paho.org/mex/index.php?option=com\\_content&view=article&id=204:antologia-atencion-salud-mexico&Itemid=315](https://www.paho.org/mex/index.php?option=com_content&view=article&id=204:antologia-atencion-salud-mexico&Itemid=315).
- PFÄFFENZELLER, S., NEWBOLD, P., AND RAYNER, A. (2007), “A Short Note on Updating the Grilli and Yang Commodity Price Index,” *The World Bank Economic Review*, **21**, 151–163.
- POSNER, D. N. (2004a), “Measuring Ethnic Fractionalization in Africa,” *American Journal of Political Science*, **48**, 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**, 529–545.
- PAMPANA, E. J. (1954), “Changing Strategy in Malaria Control,” *Bulletin of the World Health Organization*, **11**, 513.
- PAMPANA, E.J. AND RUSSELL, P.F. (1955), “Malaria: A World Problem,” *World Health Organization Chronicle*, **31**, 31.
- POWELL, R. (2006), “War as a Commitment Problem,” *International Organization*, **60**, 169–203.
- RABELL, C. A., MIER, M. AND ROCHA, T. (1986), “El descenso de la mortalidad en México de 1940 a 1980,” *Estudios demográficos y urbanos*, **1**, 39–72.
- RAMANKUTTY, N., FOLEY, J. A., NORMAN, J. *et al.* (2002), “The Global Distribution of Cultivable Lands: Current Patterns and Sensitivity to Possible Climate Change,” *Global Ecology and Biogeography*, **11**, 377–392.
- RAMOS-TORO, D. (2018), *Historical Conflict and Gender Disparities* (Brown University).
- RICHARDS, P. (1996), *Fighting for the Rain Forest: War, Youth and Resources in Sierra Leone* (Oxford, UK: The International African Institute in association with James Currey).
- ROSS, M. (2006), “A Closer Look at Oil, Diamonds and Civil War,” *Annual Review of Political Science*, **9**, 265–300.
- ROSS, M. (2008), “Oil, Islam and Women,” *American Political Science Review*, **102**, 107–123.
- RUSTAD, S. A. AND BINNINGSBØ, H. M. (2012), A Price Worth Fighting for? Natural Resources and Conflict Recurrence. *Journal of Peace Research*, **49**, 531–546.
- SACHS, J. D. AND WARNER, A. M. (1999), “The Big Push, Natural Resource Booms and Growth,” *Journal of Development Economics*, **59**, 43–76.
- SARKEES, M. AND WAYMAN, F. (2010), *Resort to War: 1816–2007* (Washington DC: CQ Press).
- SCHEFFRAN, J., BRZOSKA, M., KOMINEK, J., *et al.* (2012), “Climate Change and Violent Conflict,” *Science*, **336**, 869–871.
- TIR, J. AND 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**, 319–339.
- SELLARS, E. AND ALIX-GARCIA, J. (2018), *Locational Fundamentals, Trade, and the Changing Urban Landscape of Mexico* (Washington, D.C.: Annual Meeting Agricultural and Applied Economics Association) <https://EconPapers.repec.org/RePEc:ags:aaea18:274238>.
- UNITED NATIONS (1949), *1948 Demographic Yearbook* (Lake Success, NY: United Nations).
- UNITED NATIONS (2015), *World Population Prospects: The 2015 Revision. Volume I, Comprehensive Tables*. (New York, NY: United Nations) [https://esa.un.org/unpd/wpp/Publications/Files/WPP2015\\_Volume\\_I\\_Comprehensive-Tables.pdf](https://esa.un.org/unpd/wpp/Publications/Files/WPP2015_Volume_I_Comprehensive-Tables.pdf), (last accessed 24 March 2017).
- URDAL, H. (2005), “People vs. Malthus: Population Pressures, Environmental Degradation and Armed Conflict Revisited,” *Journal of Peace Research*, **42**, 417–434.
- URDAL, H. (2006), “A Clash of Generations? Youth Bulges and Political Violence,” *International Studies Quarterly*, **50**, 607–629.
- URLANIS, B. (2003), *Wars and Population* (Honolulu, Hawaii: University Press of the Pacific).
- WOOLDRIDGE, J. (2002), *Econometric Analysis of Cross Section and Panel Data* (Cambridge, MA: MIT Press).