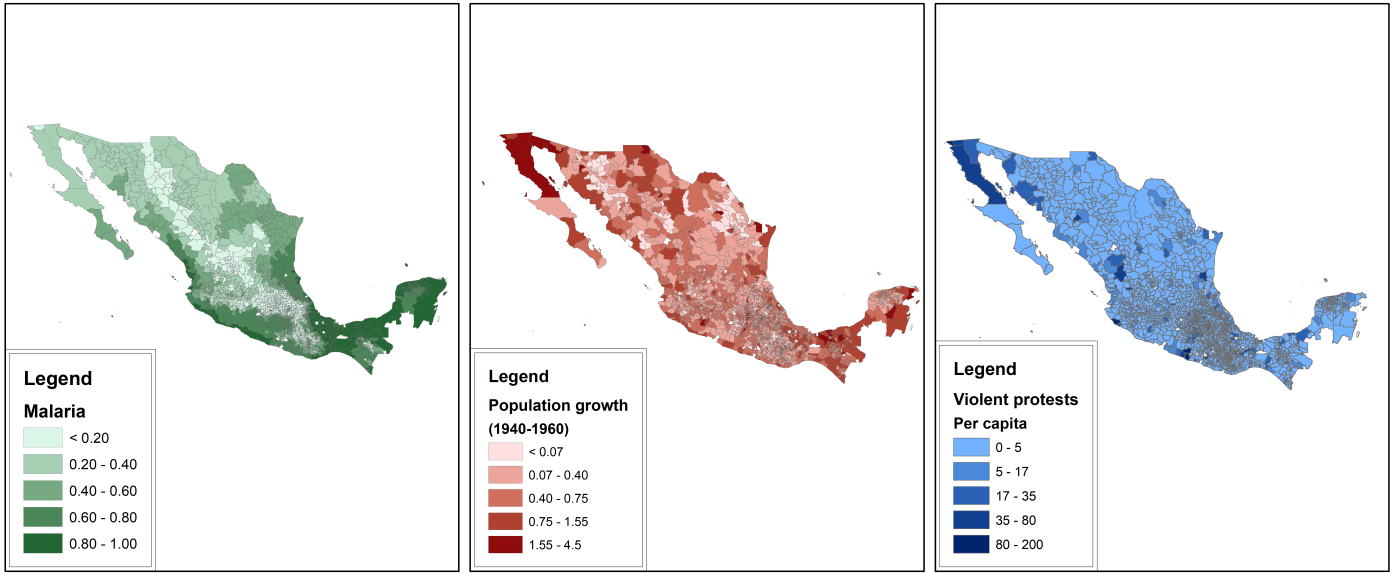
Population and Conflict
Online Appendix

Figure A-1: Malaria suitability, Population Growth and Violent Protests in Mexico



Notes: Municipal variation in malaria suitability, population growth and protest in Mexico. Malaria is temperature suitability from not suitable (0) to maximum suitability in Mexico (1). Population growth is the change in log population between 1940 and 1960. Violent protests are per capita counts (per 100.000 inhabitants) of news stories about violent protests during the 1960s. See Appendix Table A-1 for variable definition and sources.

Table A-1: Variables and Sources

Variable	Description	Source
	CROSS-COUNTRY PANEL	
	Social Conflict	
Fraction of decade in conflict	Ratio of the number of years with an internal conflict to total years assigned to reference date. Assignment of years to reference dates and exact definition of internal conflict varies by data source, as detailed below.	
COW	Number of years with intra-state war (wars that predominantly take place within the recognized territory of a state). These wars include civil wars for central control (type 4 in the COW typology) or over local issues (type 5), as well as regional internal (type 6) and intercommunal (type 7) wars. Each war in the dataset may list more than one participating country. For example, the "Overthrow of Abd el-Aziz" involves Morocco and France, and the "First Lebanese" war involves Lebanon and the US. Despite French and American involvement, we take these as civil wars in Morocco and Lebanon, as fighting took place in their territory. The threshold for inclusion in the dataset is 1,000 battle-related deaths per year (twelve-month period beginning with the start date of the war) among all the qualified war participants, including deaths from combat wounds or from diseases contracted in the war theater. Similar definition for inter-state wars, defined as those in which a member of the "interstate system" is engaged in a war with another member. It must have: sustained combat between regular armed forces on both sides and 1,000 battle-related fatalities. A war participant qualifies through either of two criteria: 100 or more fatalities or 1,000 or more armed personnel engaged in active combat. Assignment to reference dates: 1900=1900-09, 1940=1940-49, 1950=1950-59 ... 1990=90-1999, 2000=2000-2007. Downloaded from `http://www.correlatesofwar.org/data-sets.htm` last accessed on March 24, 2017.	Intra-State War Data (v4.0) and Inter-State War Data (v4.0), Correlates of War (COW). Sarkees and Wayman (2010).
UCDP/PRIO	Number of years with any incidence of an "internal armed conflict" or of "internationalized internal armed conflict". Armed conflict is defined to include all contested incompatibilities that concern government or territory or both where the use of armed force between two parties results in at least 25 battle-related deaths. Of the two parties, at least one is the government of a state. Similar definition for inter-state wars, with conflict between two or more governments – the primary warring parties must be government parties for a conflict to be classified as interstate. Assignment to reference dates: 1940=1946-49, 1950=1950-59 ... 1990=1990-99, 2000=2000-2008. Downloaded with PRIO's Battle Deaths Dataset 3.0 (see below).	UCDP/PRIO Armed Conflict Dataset, Version 4-2008. Geldtisch et. al (2002).
Fearon and Laitin	Number of years with violent civil conflicts that: (1) involved fighting between agents of (or claimants to) a state and organized, nonstate groups who sought either to take control of a government, to take power in a region, or to use violence to change government policies, (2) killed at least 1,000 over its course, with a yearly average of at least 100, (3) At least 100 were killed on both sides (including civilians attacked by rebels). Counts anticolonial wars as occurring within the empire in question (e.g., Algeria is assigned to France). Assignment to reference dates: : 1940=1945-49, 1950=1950-59 ... 1980=1980-89, 1990=1990-99.	Fearon and Laitin (2003)
Natural-resource and non-resource conflicts	Ratio of number of years with natural-resource (non-resource) conflict to total years assigned to reference date. Conflicts are coded as natural-resource related if the conflict is over the distribution of natural resources or if natural resources aggravate or finance the conflict. Based on the UCDP/PRIO conflict dataset.	Siri & Binningsbø (2012)

Continued on next page

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

Variable	Description	Source
log(1+battle deaths/ pop. in 1940)	"Best estimate" of annual battle-related deaths for use with UCDP/PRIO dataset. Assignment to reference dates as in UCDP/PRIO dataset. Population in 1940 from Maddison (2006), see below. Downloaded from `http://www.prio.org/Data/Armed-Conflict/Battle-Deaths/The-Battle-Deaths-Dataset-version-30/`, last accessed March 24, 2017.	PRIO Battle Deaths Dataset version 3.0. Lacina and Gleditsch (2005)
State Failure	Number of years where there is any serious instability. Four types of political instability are included: ethnic wars, revolutionary wars, genocides and politicides and adverse regime changes. Assignment to reference dates: 1950=1955-59, 1960=1960-69 ... 1990=1990-99, 2000=2000-2009.	Political Instability Task Force (PITF), PITF Consolidated Problem Set, Version 2013
Irregular entry and exit	Ratio of number of years with political leaders reaching power (leaving power) through irregular means to total years assigned to reference date.	Archigos, Version 2.9 (2009)
Successful and attempted coups	Ratio of number of years with successful (attempted) coups d'état to total years assigned to reference date.	Center for Systemic Peace, *Coups d'Etat, 1946-2017.*
	Population	
log population	log total population per country in 1900, 1940, 1950, 1960, 1970, 1980, 1990, 1990, 2000.	Maddison (2006)
Share of population 15-34 and 20-39	Percentage of the population ages 15-34 and 20-39 for 1940, 1950, 1960, 1970, 1980, 1990, 2000.	1950-1980: UN demographic database (`https://unstats.un.org/unsd/demographic/`). 1940: UN Demographic Yearbook 1948 (United Nations 1949, Table 4, pp. 108-158).
	Health	
Predicted mortality instrument	Sum of country's initial (in 1940) mortality rate from 14 diseases until there is a global intervention, and after the global intervention, the mortality rate from the disease in question declines to the frontier mortality rate. See paper for mathematical formula. The 14 diseases are (in rough descending order of importance): malaria, pneumonia and tuberculosis; influenza, cholera, typhoid, smallpox, whooping cough, measles (rubeola), dyphteria, scarlet fever, plague, typhus, yellow fever.	Acemoglu and Johnson (2007)
	Others	
Base Sample	Includes Acemoglu and Johnson's (2007) list of 47 non-Eastern Europe countries (Argentina, Australia, Austria, Bangladesh, Belgium, Brazil, Canada, Chile, China, Colombia, Costa Rica, Denmark, Ecuador, El Salvador, Finland, France, Germany, Greece, Guatemala, Honduras, India, Indonesia, Ireland, Italy, South Korea, Rep., Malaysia, Mexico, Myanmar, Netherlands, New Zealand, Nicaragua, Norway, Pakistan, Panama, Paraguay, Peru, Philippines, Portugal, Spain, Sri Lanka, Sweden, Switzerland, Thailand, United Kingdom, United States, Uruguay, Venezuela); the set of 12 additional countries for which they have life expectancy data since 1950 (Algeria, Bolivia, Egypt, Iran, Iraq, Lebanon, Morocco, Singapore, South Africa, Tunisia, Turkey and Vietnam); and 6 countries from Eastern Europe (Bulgaria, Czech Republic, Hungary, Poland, Romania and Russian Federation). This implies a total of 65 countries, but not all have all variables for all years. In particular, 13 countries lack population data and/or had not yet been created in 1940 (Algeria, Bangladesh, Egypt, Iran, Iraq, Lebanon, Malaysia, Morocco, Russia, Singapore, South Africa, Tunisia and Vietnam). Also, Austria is excluded in 1940 when the dependent variables are from COW since it enters the COW state system in the 1950s.	

Continued on next page

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

Variable	Description	Source
Initially rich, middle-income, and poor countries	Each category is defined using the top, middle and lowest third group of countries in the base sample based on income per capita in 1940. Initially rich countries had log GDP per capita over 8.4. in 1940; middle income had log GDP per capita betweeen 7.37 and 8.4; and low income countries had log GDP per capita below 7.37 in 1940.	
Country clusters	In computing standard errors, Bangladesh, India and Pakistan are considered a single cluster.	
Countries most affected by World War II	Austria, China, Finland, Germany, Italy, Russian Federation	Acemoglu and Johnson (2007) based on Urlanis (2003)
	Controls and Baseline characteristics	
Agriculture, value added	Value added in agriculture as percentage of GDP	World Development Indicators, World Bank
Agricultural suitability	*Climate:* Index of land suitability for cultivation based on climate (temperature, precipitation and potential sunshine hours). *Soil*: Index of land suitability for cultivation based on soil characteristics (soil carbon density and soil pH). *Average*: Index of land suitability for cultivation that indicates whether the characteristics of the land allow for cultivation. The overall index is the product of two components (climate and soil). See `https://assets.aeaweb.org/assets/production/articles-attachments/aer/data/june2012/20090179_app.pdf` for details on the exact formulas used in the construction of the land quality index.	Michalopoulos (2012) and Ramankutty et al. (2002)
Average constraints on the executive	Average of constraints on the executive in the 1940s.	Polity IV
Average polity score	Average polity score in the 1940s.	Polity IV
Cereal suitability index	Reflects crop suitability levels, between 0 and 100, of cereals based on the soil, the agro-climatically attainable biomass and the potential productivity estimates.	FAO/IIASA (2010). Data from the GAEZ data portal of FAO (`http://gaez.fao.org/Main.html`).
Commodity price shocks	$P_{it} = \sum_{j=1}^{J} \Delta p_{jt} \times \omega_{ji}$, where Δp_{jt} is the yearly change in the price of commodity j and ω_{ji} is the export per capita of commodity j in country i in the 1960s. Commodities included are coffee, cocoa, tea, rice, wheat, maize, sugar, beef, lamb, banana, palmoil, cotton, jute, wool, hides, tobacco, rubber, timber, copper, aluminum, tin, silver, lead, and zinc.	Prices: Grilli & Yang (1988), updated by Pfaffenzeller et al. (2007); Export weights: UN Comtrade dataset.
Country area	Country's area in square kilometers.	Weidmann & Gleditsch (2016)
Educational attainment	Average educational attainment (years of schooling) for population aged 15 and over.	Barro & Lee (2013)
Ethnic and religious composition and polarization	Ethnic Polarization; Ethnic Fragmentation; Religious Polarization; Religious Fragmentation Ethnolinguistic fractionalization index (from 0 to 1). Average value of five indices based on ethnic or linguistic characteristics of the population. Share of Muslim, Catholic and Protestant Populations in 1980	Montalvo and Reynal-Querol (2005) Easterly and Levine (1997) La Porta et al (1999).

Continued on next page

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

Variable	Description	Source
Ethnic Dominance	Dummy variable that takes the value of 1 if one ethnolinguistic group represented between 45% and 90% of the population in 1960	Montalvo and Reynal-Querol (2005)
Ethnic Inequality	Gini coefficients capturing the differences in mean income - as captured by luminosity per capita of the ethnic homeland - across ethnolinguistic groups. The difference between the two measures is the source of the data used to identify the location of the ethnic groups. The first measure uses the Georeferencing of Ethnic Groups (GREG) while the second uses the 15th edition of the Ethnologue (Gordon, 2005).	Alesina et al. (2014)
Institutions	Average of constraints on the executive in 1950, 1960 and 1970	Polity IV
Independent in 1940	=1 if country is independent in 1940, 0 otherwise	Own coding.
Initial GDP	log GDP per capita in 1930	Maddison (2006)
Latitude	Latitude of the centroid of the country.	Portland State University Country Geography Data (`https://www.pdx.edu/econ/country-geography-data%20/`)
Malaria ecology index	Index of malaria transmission based on climate (rainfall and temperature) and the dominant properties of anopheline vectors of malaria.	Kiszewski et al. (2004). Downloaded from `https://sites.google.com/site/gordoncmccord/datasets`
Natural resources	Oil production and oil production per capita in 1960; Diamond production and diamond production per capita in 1960 Share of natural resource sector in GNP in 1970; Share of mineral production in GNP in 1971 Oil and gas rents per capita in 1960	Humphreys (2005) Sachs and Warner (1999) Ross (2008)
Oil gini	Gini coefficients capturing the unevenness of oil field distribution across ethnic groups for a given country	Morelli & Rohner (2015)
Population density	Population per square kilometer.	Own coding.
US and Soviet Influence	US Influence is a yearly dummy variable that equals one if the CIA either installed a foreign leader or provided covert support for the regime once in power. Soviet influence is a similar indicator variable for KGB interventions defined in the same manner as the CIA interventions. From these indicators we count the fraction of years from 1940 to 1980 with US or Soviet Influence.	Berger, Easterly, Nunn, & Satyanath (2013)

Continued on next page

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

Variable	Description	Source
	WITHIN-COUNTRY (MEXICO)	
	Social Conflict	
Protests	Per capita counts (per 100.000 people in 1940) of news stories about protests published from 1960 to 1969 in newspapers *Excelsior* and *El Universal*. We proceed in two steps. In the first step, we manually identified in the two newspapers from 1960 to 1969 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⋆". Next, we identify the news stories in which the title or description is found to be related to violence, conflict, arms, social disorder, or aggression. Specifically, we use catch words to code news stories as related to *violent protests*. These catch words are: "agita", "desorden", "violencia", "violación", "armado", "agresión" and "conflicto". To measure *non-violent protests* we simply count all news stories that do not have these catch words for violence. *Natural-resource and Non-resource violent protests:* Natural-resource protests are violent protests in which the title or description of the news story mentions natural resources such as water, land, mining and agricultural goods. The rest are non-resource violent protest.	Fergusson, Larreguy & Riaño (2018)
Historical conflicts	Episodes of armed conflict from 1616 to 1940, georeferenced at the municipal level. Where only the state is known, we attribute the episode proportionally to all municipalities within the state.	Ramos-Toro (2018)
	Population	
log population	log population at the municipal level.	INEGI, Censo de población 1940 and Censo de población 1960 and Sellars and Alix-Garcia (2018)
Share of pop. i-j	Percentage of the population ages i to j.	INEGI, Censo de población 1940 and Censo de población 1960
	Health	
Predicted mortality (Mexico)	In 1940, index of temperature suitability for the transmission of malaria, between 0 (unsuitable) to 1 (maximum suitability in Mexico). The index for each municipality is the average malaria suitability within its boundaries from a 1 km^2 resolution land pixels.	Gething et al. (2011)
	Others	
Droughts	Number of months with droughts in the 1960s. A drought is defined as precipitation below the 5th percentile of the long-run distribution (1900-2008) of monthly rain per municipality. Computed separately for the non-harvesting (January to September) and harvesting period (October to December) for corn.	Matsuura & Willmott (2009), University of Delaware. Downloaded from `http://research.jisao.washington.edu/data/ud/`

Continued on next page

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

Variable	Description	Source
Literacy	Literacy rate of population aged 6 years and older.	INEGI, Censo de población 1940 and Censo de población 1960
Literacy 25-39	Literacy rate for ages 25-39.	INEGI, Censo de población 1940 and Censo de población 1960
Primary School, 1940	Percentage of the population older than 12 years having completed primary school in 1940	INEGI, Censo de población 1940 and Censo de población 1960
University enrollment, 1940	Share of population enrolled in a university level program.	INEGI, Censo de población 1940
Distance to capital	Linear distance from a municipality's centroid to Mexico City.	Own coding.
Distance to large cities	Linear distance from a municipality's centroid to the nearest municipality with a total population of at least 100,000 in 1960.	Own coding.
Land quality	Soil resilience and performance index, ranging from 1 to 9 (higher values indicate higher land quality), based on climate and geological factors. The index for each municipality is the average land quality within its boundaries.	Fergusson, Larreguy & Riaño (2018)
Sedimentary basin	Share of a municipality's area on a sedimentary basin – areas where petroleum can potentially form. Following Cassidy (2018), we use the following categories of the plate-tectonic environment: Convergent C-C, Convergent O-C, Divergent and Wrench (where "C" stands for continent and "O" stands for "Ocean"). Note that we exclude Convergent O-O because in Cassidy (2018) it has a negative relationship with the annual oil production per capita at the cross-country level.	Data from the Fugro Robertson, Ltd. (2013) Tellus GIS database.
Share indigenous	Percentage of the population speaking only indigenous languages.	INEGI, Censo de población 1940.

Table A-2: Population and Conflict: Additional Robustness Wild Bootstrap Procedure and Global Mortality Instrument

	Dependent variable is fraction of decade in conflict according to Correlates of War (cols 1, 2 and 4) & log population (col 3)			
	Wild Bootstrap Procedure		Global Mortality Instrument	
	OLS	2SLS	First Stage	2SLS
	(1)	(2)	(3)	(4)
Panel A: Long differences, just 1940s and 1980s				
log population	0.323	0.617		0.624
	[0.018]	[0.016]		(0.216)
Global predicted mortality			-0.818	
			(0.207)	
Observations	102	102	102	102
R-squared	0.627	0.572	0.823	0.049
Number of clusters	50	50	50	50
Panel B: Panel regressions, 1950s-1980s				
log population	0.268	0.609		0.570
	[0.002]	[0.010]		(0.191)
Global predicted mortality			-0.681	
			(0.144)	
Observations	307	307	307	307
R-squared	0.482	0.430	0.814	0.014
Number of clusters	63	63	63	63

Notes: Columns 1 and 3: OLS regressions with a full set of year and country fixed effects (column 1 is equation (4.4) in the text, column 3 is equation (4.6) in the text). Columns 2 and 4: 2SLS regressions with a full set of year and country fixed effects (equation (4.4) in the text, where population is instrumented with the baseline predicted mortality instrument in column 2, and with the global mortality instrument in column 4, as in equation (4.6) in the text). Robust standard errors (clustered by country) are reported in parentheses. Cluster robust p-values following the wild bootstrap procedure suggested by Cameron, Gelbach and Miller (2008) are reported in square brackets. 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. In computing standard errors, Bangladesh, India and Pakistan are considered a single cluster. See the text and Appendix Table A-1 for definitions and details.

Table A-3: Population and Conflict: OLS Estimates
Share of Population 20-39

Dependent variable...	Fraction of decade in conflict			
	COW	UCDP/PRIO	Fearon & Laitin	log(1+battle deaths/ pop.1940)
	(1)	(2)	(3)	(4)
Panel A: Long differences controlling for age structure, just 1940s and 1980s				
log population	0.377	0.321	0.318	0.989
	(0.136)	(0.148)	(0.154)	(0.458)
Share of population 20-39	-1.155	-1.741	-5.836	-6.341
	(1.467)	(1.574)	(3.019)	(4.729)
Observations	86	88	88	88
Number of clusters	43	44	44	44
Panel B: Panel regressions controlling for age structure, 1940s-1980s				
log population	0.316	0.301	0.243	0.917
	(0.123)	(0.151)	(0.145)	(0.434)
Share of population 20-39	-0.861	-1.127	-2.430	-3.716
	(0.600)	(0.670)	(1.623)	(1.963)
Observations	227	228	228	228
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. Panel A are long-difference specifications with two observations per country, one for the initial date and one for the final date. Panel B 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 Appendix Table A-1 for definitions and details.

Table A-4: OLS and 2SLS Estimates
Spatial Correction of Standard Errors

Dependent variable is fraction of decade in conflict according to Correlates of War		
	(1)	(2)
	OLS	2SLS
Panel A: Long differences, just 1940s and 1980s		
log population	0.323	0.617
	(0.035)	(0.104)
Observations	102	102
Panel B: Panel regressions, 1940s-1980s		
log population	0.268	0.609
	(0.050)	(0.232)
Observations	307	307

Notes: Column 1 are OLS regressions with a full set of year and country fixed effects (equation (4.4) in the text). Column 2 are 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 reported in parentheses allow for spatial correlation within a maximum radius of 9.684km, which equals the average distance between any pair of countries in our sample. 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. See the text and Appendix Table A-1 for definitions and details.

Table A-5: OLS and 2SLS Estimates
Inverse Hyperbolic Sine and Linear Probability Model

	Dependent variable is ...			
	Inverse hyperbolic sine of battle deaths to initial (1940) population		Dummy=1 if any civil war in decade	
	(1)	(2)	(3)	(4)
	OLS	2SLS	OLS	2SLS
Panel A: Long differences, just 1940s and 1980s				
log population	0.886	1.724	0.312	1.093
	(0.502)	(0.745)	(0.196)	(0.309)
Observations	102	102	102	102
Number of clusters	50	50	50	50
Panel B: Panel regressions, 1940s-1980s				
log population	0.895	1.404	0.356	1.234
	(0.461)	(0.573)	(0.177)	(0.431)
Observations	273	273	307	307
Number of clusters	54	54	63	63

Notes: Column 1 and 3 are OLS regressions with a full set of year and country fixed effects (equation (4.4) in the text). Column 2 and 4 are 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. Bangladesh, India and Pakistan are considered a single cluster. See the text and Appendix Table A-1 for definitions and details.

Table A-6: The Effect of Population on Political Instability and Inter-State Conflict: 2SLS Estimates

	Dependent variable is fraction of decade with...						
						Inter-State conflicts in...	
	Irregular entry	Irregular exit	Successful coups	Attempted coups	State failure	COW	UCDP/PRIO
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Long differences, just 1940s and 1980s (for state failure, using 1950s as 1940s)							
log population	0.003	0.000	-0.068	0.090	0.743	0.114	-0.084
	(0.052)	(0.065)	(0.096)	(0.112)	(0.209)	(0.306)	(0.089)
Observations	102	102	98	98	102	102	104
Number of clusters	50	50	49	49	50	50	51
Panel B: Panel regressions, 1940s-1980s (for state failure, 1950s-1980s)							
log population	0.050	0.018	-0.005	0.009	1.649	0.172	-0.069
	(0.067)	(0.079)	(0.150)	(0.117)	(0.689)	(0.490)	(0.188)
Observations	314	314	303	303	260	307	308
Number of clusters	63	63	62	62	63	63	63
Dep var mean	0.03	0.03	0.02	0.03	0.15	0.075	0.041
Dep var s.d.	0.09	0.08	0.08	0.08	0.31	0.18	0.13

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. In computing standard errors, Bangladesh, India and Pakistan are considered a single cluster. See the text and Appendix Table A-1 for definitions and details.

Table A-7: Robustness to Sample Selection: Asia, Africa, America & Australia

	Dependent variable is log population (cols 1-4) & fraction of decade in conflict according to Correlates of War (cols 5-12)											
	First Stages (cols 1-4)				Reduced Forms (cols 5-8)				2SLS estimates (cols 9-12)			
Excluding...	Asia	Africa	America	Australia	Asia	Africa	America	Australia	Asia	Africa	America	Australia
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Long differences, just 1940s and 1980s												
Baseline predicted mortality	-0.920	-0.782	-0.661	-0.813	-0.436	-0.482	-0.412	-0.491				
	(0.189)	(0.141)	(0.157)	(0.144)	(0.209)	(0.178)	(0.238)	(0.181)				
log population									0.474	0.617	0.624	0.604
									(0.227)	(0.213)	(0.300)	(0.208)
Observations	82	102	64	98	82	102	64	98	82	102	64	98
Number of clusters	41	50	31	48	41	50	31	48	41	50	31	48
Panel B: Panel regressions, 1940s-1980s												
Baseline predicted mortality	-0.544	-0.463	-0.408	-0.485	-0.154	-0.301	-0.357	-0.298				
	(0.126)	(0.094)	(0.101)	(0.097)	(0.074)	(0.089)	(0.141)	(0.090)				
log population									0.299	0.610	0.830	0.599
									(0.137)	(0.206)	(0.329)	(0.200)
Observations	232	288	212	297	235	292	218	303	232	288	212	297
Number of clusters	48	58	44	61	48	58	44	61	48	58	44	61

Notes: Columns 1-4 are OLS regressions with a full set of year and country fixed effects (equation (6) in the text). Columns 5-8 are OLS regressions with a full set of year and country fixed effects (equation (4) in the text, using predicted mortality instead of log population as regressor). Columns 9-12 are 2SLS regressions with a full set of year and country fixed effects (equation (4) in the text, where population is instrumented by predicted mortality, as in equation (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. In computing standard errors, Bangladesh, India and Pakistan are considered a single cluster. See the text and Appendix Table A-1 for definitions and details.

Table A-8: Conflict and Predicted Mortality
Reduced Form Basic Robustness

	Dependent variable is fraction of decade in conflict according to Correlates of War								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Base Sample	Excluding Eastern Europe	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	Just 1940s and 2000s
Panel A: Long differences, just 1940s and 1980s									
Baseline predicted mortality	-0.482	-0.460	-0.575	-0.632	-0.495	-0.328	-0.500	-0.597	-0.347
	(0.178)	(0.176)	(0.224)	(0.234)	(0.188)	(0.167)	(0.185)	(0.213)	(0.153)
Observations	102	92	72	80	94	102	102	86	102
Number of clusters	50	45	35	39	46	50	50	43	50
Panel B: Panel regressions, 1940s-1980s									
Baseline predicted mortality	-0.291	-0.261	-0.350	-0.394	-0.292	-0.131	-0.314	-0.254	-0.315
	(0.088)	(0.084)	(0.113)	(0.124)	(0.092)	(0.053)	(0.100)	(0.104)	(0.087)
Observations	313	283	234	258	284	313	313	227	443
Number of clusters	63	57	47	52	57	63	63	46	63

Notes: OLS regressions with a full set of year and country fixed effects (regressions as in equation (4.4) in the text, using predicted mortality instead of log population as a regressor). Robust standard errors (clustered by country) are reported in parentheses. Panel A presents long-difference specifications with two observations per country, one for the initial date and one for the final date. Panel B presents unbalanced panels with one observation per decade. In computing standard errors, Bangladesh, India and Pakistan are considered a single cluster. See the text and Appendix Table A-1 for definitions and details.

Table A-9: Predicted Mortality and Age Structure

	Dependent variable is share of population ages 15-34 and 20-39			
	(1)	(2)	(3)	(4)
	Base Sample	Low and Middle Income Countries Only	Base Sample	Low and Middle Income Countries Only
	Long differences, just 1940s and 1980s		Long differences, just 1940s and 1990s	
Panel A: Dependent variable is share of population ages 15-34				
Baseline predicted mortality	0.004	-0.000	-0.028	-0.030
	(0.017)	(0.028)	(0.019)	(0.035)
Observations	86	64	86	64
R-squared	0.017	0.025	0.057	0.038
Number of clusters	43	32	43	32
Panel B: Dependent variable is share of population ages 20-39				
Baseline predicted mortality	0.016	0.018	-0.002	-0.003
	(0.014)	(0.021)	(0.019)	(0.031)
Observations	86	64	86	64
R-squared	0.281	0.314	0.001	0.001
Number of clusters	43	32	43	32

Notes: OLS regressions with a full set of year and country fixed effects (regressions as in equation (4.4) in the text, using share of population ages 15-34 instead of conflict as dependent variable). Robust standard errors (clustered by country) are reported in parentheses. In computing standard errors, Bangladesh, India and Pakistan are considered a single cluster. See the text and Appendix Table A-1 for definitions and details.

Table A-10: Timing of the Effect of Population on Conflict: 2SLS Estimates

	Dependent variable is fraction of decade in conflict according to Correlates of War					
	(1)	(2)	(3)	(4)	(5)	(6)
						Interaction with US or Soviet influence
Panel A: Long differences	Just 1940s and 1960s	Just 1940s and 1970s	Just 1940s and 1980s	Just 1940s and 1990s	Just 1940s and 2000s	Just 1940s and 1980s
log population	0.389	0.600	0.617	0.409	0.296	0.551
	(0.304)	(0.237)	(0.213)	(0.164)	(0.144)	(0.172)
log population × US or Soviet influence						0.405
						(0.189)
Observations	102	102	102	102	102	100
Number of clusters	50	50	50	50	50	49
Panel A: Long differences	1940s–1960s	1940s–1970	1940s–1980s	1940s–1990s	1940s–2000s	1940s–1980s
log population	0.641	0.615	0.609	0.539	0.467	0.548
	(0.384)	(0.272)	(0.205)	(0.167)	(0.152)	(0.182)
log population × US or Soviet influence						0.329
						(0.164)
Observations	175	241	307	372	437	298
Number of clusters	61	63	63	63	63	61

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. Long-difference specifications with two observations per country, one for the initial date and one for the final date. In computing standard errors, Bangladesh, India and Pakistan are considered a single cluster. See the text and Appendix Table A-1 for definitions and details.

Table A-11: Conflict and Predicted Mortality
Reduced Form Robustness to Differential Trends

	Dependent variable is fraction of decade in conflict according to Correlates of War									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
				Interaction of post-year dummies with...						
	Institutions	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
Panel A: Long differences, just 1940s and 1980s										
Baseline predicted mortality	-0.546	-0.679	-0.470	-0.603	-0.483	-0.482	-0.442	-0.465	-0.467	-0.515
	(0.204)	(0.242)	(0.166)	(0.211)	(0.179)	(0.178)	(0.192)	(0.209)	(0.187)	(0.190)
Observations	102	100	102	86	102	102	96	88	102	102
Number of clusters	50	49	50	43	50	50	47	44	50	50
p-value for post year dummy x variable indicated at the top of each column	0.142	0.149	0.0875	0.445	0.149	0.863	0.350	0.462	0.699	0.302
Panel B: Panel regressions,1940s-1980s										
Baseline predicted mortality	-0.316	-0.446	-0.275	-0.256	-0.292	-0.283	-0.235	-0.199	-0.261	-0.299
	(0.105)	(0.135)	(0.091)	(0.105)	(0.088)	(0.087)	(0.091)	(0.077)	(0.087)	(0.085)
Observations	313	268	265	223	305	305	281	249	305	313
Number of clusters	63	53	52	45	61	61	56	50	61	63
p-value for post year dummy x variable indicated at the top of each column	0.561	0.0788	0.0252	0.733	<0.01	0.735	0.292	0.103	0.523	0.180

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

Table A-12: Population and Predicted Mortality
First and Second Stage Robustness to Additional Differential Trends

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
					Interaction of post-year dummies with...						
	Share of GDP in Natural Resource Sector in 1970	Share of Mineral Production in GNP in 1971	Oil Production in 1960	Diamond Production in 1960	Religious Polarization	Religious Fragmentation	Average Ethno-linguistic Fragmentation	Ethnic Fragmentation	Share Catholic Muslim Protestant in 1980	Independent in 1940	GDP per capita 1960 terciles
Panel A: 2SLS estimates. Dependent variable is fraction of decade in conflict (Long differences, just 1940s and 1980s)											
log population	1.415	0.720	0.605	0.627	0.848	0.772	0.555	0.737	0.965	0.563	0.985
	(0.558)	(0.265)	(0.206)	(0.214)	(0.371)	(0.288)	(0.275)	(0.326)	(0.380)	(0.220)	(0.492)
p-value for post year dummy x variable indicated at the top of each column	0.0553	0.0632	<0.01	0.0358	0.365	0.270	0.678	0.600	0.161	0.515	<0.01
Panel B: First stage. Dependent variable is log population (Long differences, just 1940s and 1980s)											
Baseline predicted mortality	-0.495	-0.661	-0.796	-0.772	-0.562	-0.636	-0.777	-0.587	-0.584	-0.815	-0.486
	(0.191)	(0.145)	(0.137)	(0.138)	(0.174)	(0.152)	(0.201)	(0.168)	(0.185)	(0.175)	(0.196)
R-squared	0.888	0.853	0.844	0.837	0.848	0.845	0.823	0.849	0.842	0.824	0.864
p-value for post year dummy x variable indicated at the top of each column	0.953	0.167	<0.01	0.0126	0.0515	0.0588	0.961	0.0605	0.158	0.582	<0.01
Observations (Panels A and B)	68	92	102	102	96	96	102	96	100	102	102
Number of clusters (Panels A and B)	33	45	50	50	47	47	50	47	49	50	50
Panel C: 2SLS estimates. Dependent variable is fraction of decade in conflict (Panel regressions, 1940s-1980s)											
log population	1.498	0.667	0.600	0.618	0.710	0.672	0.356	0.569	0.945	0.497	0.857
	(0.668)	(0.259)	(0.203)	(0.205)	(0.353)	(0.275)	(0.181)	(0.270)	(0.410)	(0.166)	(0.461)
p-value for post year dummy x variable indicated at the top of each column	0.165	0.0166	0.0804	<0.01	0.701	0.472	0.147	0.111	0.441	0.0707	0.176
Panel D: First stage. Dependent variable is log population (Panel regressions, 1940s-1980s)											
Baseline predicted mortality	-0.268	-0.379	-0.469	-0.455	-0.312	-0.362	-0.429	-0.326	-0.359	-0.468	-0.284
	(0.112)	(0.092)	(0.089)	(0.090)	(0.107)	(0.096)	(0.116)	(0.099)	(0.123)	(0.114)	(0.123)
R-squared	0.871	0.835	0.813	0.801	0.839	0.829	0.804	0.837	0.834	0.799	0.860
p-value for post year dummy x variable indicated at the top of each column	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.0113	<0.01
Observations (Panels C and D)	197	273	306	306	283	283	298	283	301	306	306
Number of clusters (Panels C and D)	38	54	61	61	56	56	59	56	60	61	61

Notes: Panels A and C: 2SLS regressions with a full set of year and country fixed effects (equation (4.9) 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 Appendix Table A-1 for definitions and details.

Table A-13: Conflict and Predicted Mortality
Reduced Form Robustness to Additional Differential Trends

	Dependent variable is fraction of decade in conflict according to Correlates of War										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Interaction of post-year dummies with...										
	Share of GDP in Natural Resource Sector in 1970	Share of Mineral Production in GNP in 1971	Oil Production in 1960	Diamond Production in 1960	Religious Polarization	Religious Fragmentation	Average Ethnolinguistic Fragmentation	Ethnic Fragmentation	Share Catholic Muslim Protestant in 1980	Independent in 1940	GDP per capita 1960 terciles
Panel A: Long differences, just 1940s and 1980s											
Baseline predicted mortality	-0.700	-0.476	-0.481	-0.484	-0.477	-0.492	-0.431	-0.433	-0.563	-0.459	-0.479
	(0.232)	(0.182)	(0.178)	(0.179)	(0.228)	(0.201)	(0.199)	(0.198)	(0.208)	(0.179)	(0.196)
Observations	68	92	102	102	96	96	102	96	100	102	102
Number of clusters	33	45	50	50	47	47	50	47	49	50	50
p-value for post year dummy x variable indicated at the top of each column	0.0319	0.331	0.551	0.115	0.942	0.924	0.660	0.539	0.212	0.738	<0.01
Panel B: Panel regressions, 1940s-1980s											
Baseline predicted mortality	-0.370	-0.247	-0.285	-0.292	-0.204	-0.231	-0.185	-0.188	-0.318	-0.244	-0.248
	(0.131)	(0.084)	(0.086)	(0.088)	(0.107)	(0.099)	(0.068)	(0.081)	(0.114)	(0.074)	(0.092)
Observations	192	271	305	305	281	281	295	281	300	305	305
Number of clusters	38	54	61	61	56	56	59	56	60	61	61
p-value for post year dummy x variable indicated at the top of each column	0.397	0.365	0.832	<0.01	0.660	0.633	0.0599	0.146	0.0460	0.0277	<0.01

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

Table A-14: Rotemberg Weights

	Base Sample	Excluding Eastern Europe	Excluding Western Europe	Low and Middle Income Countries Only	Excluding Most Affected By WWII	Excluding Pneumonia
Panel A: Long differences, just 1940s and 1980s						
Diphtheria, 1940	-0.008	-0.006	-0.006	-0.009	-0.006	-0.017
Scarlet fever, 1940	-0.002	-0.002	-0.003	-0.003	-0.002	-0.005
Whooping cough, 1940	0.003	0.005	0.005	0.002	0.003	0.008
Smallpox, 1940	0.005	0.003	0.001	0.003	0.004	0.012
Plague, 1940	0.009	0.009	0.011	0.012	0.009	0.021
Cholera, 1940	0.010	0.008	-0.001	0.009	0.008	0.022
Measles (rubeola), 1940	0.023	0.021	0.033	0.028	0.021	0.051
Typhus, 1940	0.027	0.027	0.024	0.032	0.026	0.061
Typhoid, 1940	0.029	0.030	0.028	0.032	0.028	0.064
Influenza, 1940	0.039	0.028	0.053	0.044	0.035	0.087
Malaria, 1940	0.126	0.117	0.091	0.142	0.117	0.284
Tuberculosis, 1940	0.184	0.190	0.227	0.102	0.235	0.413
Pneumonia, 1940	0.555	0.570	0.538	0.606	0.522	0
Panel B: Panel regressions, 1940s-1980s						
Diphtheria, 1940	-0.009	-0.007	-0.008	-0.010	-0.007	-0.021
Scarlet fever, 1940	-0.002	-0.002	-0.003	-0.004	-0.003	-0.006
Smallpox, 1950	0.002	0.001	-0.002	0.000	0.001	0.005
Whooping cough, 1940	0.003	0.006	0.005	0.002	0.002	0.008
Smallpox, 1940	0.004	0.002	-0.002	0.002	0.004	0.010
Cholera, 1940	0.007	0.004	-0.008	0.004	0.005	0.016
Plague, 1940	0.011	0.011	0.014	0.014	0.011	0.026
Measles (rubeola), 1940	0.024	0.023	0.036	0.030	0.022	0.057
Typhus, 1940	0.029	0.028	0.026	0.034	0.028	0.067
Typhoid, 1940	0.031	0.034	0.033	0.035	0.030	0.073
Influenza, 1940	0.037	0.023	0.052	0.041	0.033	0.087
Malaria, 1940	0.110	0.096	0.056	0.117	0.100	0.260
Tuberculosis, 1940	0.177	0.184	0.221	0.101	0.229	0.417
Pneumonia, 1940	0.575	0.597	0.579	0.632	0.545	0

Notes: The table presents Rotemberg weights for the diseases (times time period) used in the construction of predicted mortality. 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 balanced panels with one observation per decade. In both panels, we report the Rotemberg weights only for diseases (times time period) with a weight different than 0 in one of our specifications. .

Table A-15: Bartik Instruments: 2SLS Estimates

	Dependent variable is fraction of decade in conflict according to Correlates of War								
Instrumental variable is...	Predicted mortality	Initial disease mortality times time period	Excluding pneumonia	Excluding malaria	Excluding tuberculosis	Excluding pneumonia, malaria and tuberculosis	Just pneumonia	Just malaria	Just tuberculosis
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Long differences, just 1940s and 1980s									
log population	0.617	0.394	0.987	0.605	0.436	0.677	0.319	0.694	1.416
	(0.213)	(0.139)	(0.340)	(0.217)	(0.181)	(0.295)	(0.170)	(0.344)	(0.794)
Observations	102	102	102	102	102	102	102	102	102
Number of clusters	50	50	50	50	50	50	50	50	50
Panel B: Panel regressions, 1940s-1980s									
log population	0.609	0.268	0.762	0.621	0.514	0.601	0.491	0.512	1.057
	(0.205)	(0.120)	(0.254)	(0.220)	(0.203)	(0.232)	(0.266)	(0.246)	(0.655)
Observations	307	307	307	307	307	307	307	307	307
Number of clusters	63	63	63	63	63	63	63	63	63

Notes: 2SLS regressions with a full set of year and country fixed effects (equation (4) in the text, where population is instrumented by the variable shown in each column title, as in equation (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 balanced panels with one observation per decade. Column 1 uses the standard predicted mortality for comparison. Column 2 uses each disease mortality rate times time period independently as instruments. Columns 3 to 5 exclude pneumonia, malaria and tuberculosis when constructing the predicted mortality instrument. Column 6 excludes these three diseases simultaneously from the instrument. And columns 7 to 9 use only the rates of pneumonia, malaria and tuberculosis as instruments. In computing standard errors, Bangladesh, India and Pakistan are considered a single cluster. See the text and Appendix Table A-1 for definitions and details.

Table A-16: Bartik Instruments: First-stage Estimates

	Dependent variable is log population							
Instrumental variable is...	Predicted mortality	Excluding pneumonia	Excluding malaria	Excluding tuberculosis	Excluding pneumonia, malaria and tuberculosis	Just pneumonia	Just malaria	Just tuberculosis
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Long differences, just 1940s and 1980s								
Variable shown in each column title	-0.782	-0.797	-0.860	-1.012	-2.473	-1.381	-1.536	-0.720
	(0.141)	(0.277)	(0.162)	(0.220)	(0.822)	(0.277)	(1.087)	(0.520)
Observations	102	102	102	102	102	102	102	102
Number of clusters	50	50	50	50	50	50	50	50
Panel B: Panel regressions, 1940s-1980s								
Variable shown in each column title	-0.464	-0.458	-0.520	-0.608	-1.550	-0.833	-0.782	-0.405
	(0.094)	(0.182)	(0.105)	(0.145)	(0.504)	(0.185)	(0.676)	(0.313)
Observations	307	307	307	307	307	307	307	307
Number of clusters	63	63	63	63	63	63	63	63

Notes: OLS regressions with a full set of year and country fixed effects (equation (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 balanced panels with one observation per decade. Column 1 uses the standard predicted mortality for comparison. Columns 2 to 4 exclude pneumonia, malaria and tuberculosis when constructing the predicted mortality instrument. Column 5 excludes these three diseases simultaneously from the instrument. And columns 6 to 8 use only the rates of pneumonia, malaria and tuberculosis as instruments. The first stage for column 2 of Appendix Table A-15 is not shown to save space. The F-stat of the excluded instruments is 68.49 (p-value < 0.0001) for the long-differences specification and 2301.33 (p-value < 0.0001) for the panel regression. In computing standard errors, Bangladesh, India and Pakistan are considered a single cluster. See the text and Appendix Table A-1 for definitions and details.

Table A-17: Population and Conflict: the Role of Education
2SLS Estimates

	Dependent variable is educational attainment (av. years of schooling)	Dependent variable is fraction of decade in conflict (cols 2-4) and log (1+ battle deaths/pop. in 1940) in col 5			
		COW	UCDP/PRIO	Fearon & Laitin	log(1+battle deaths/ pop. 1940)
	(1)	(2)	(3)	(4)	(5)
Panel A: Long differences, just 1940s and 1980s (for educational attainment using 1950s as 1940s)					
log population	-0.234	0.621	0.558	0.852	1.310
	(0.664)	(0.214)	(0.224)	(0.256)	(0.584)
Educational attainment		0.018	-0.091	-0.142	-0.190
		(0.053)	(0.057)	(0.073)	(0.130)
Observations	102	102	104	104	104
Number of clusters	50	50	51	51	51
Panel B: Panel regressions, 1940s-1980s (for educational attainment using 1950s as 1940s)					
log population	-0.232	0.601	0.285	0.858	1.077
	(0.550)	(0.208)	(0.232)	(0.431)	(0.432)
Educational attainment		-0.032	-0.094	-0.081	-0.156
		(0.028)	(0.036)	(0.037)	(0.078)
Observations	303	303	304	304	273
Number of clusters	62	62	62	62	54

Notes: Column 1: 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). Columns 2-5: 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. In computing standard errors, Bangladesh, India and Pakistan are considered a single cluster. See the text and Appendix Table A-1 for definitions and details.

Table A-18: The Effect of Population on Conflict: 2SLS Estimates
Heterogeneous Effects I

	Dependent variable is fraction of decade in conflict according to Correlates of War									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		Interactions of log population with...								
	GDP per capita	Ethnic Polarization	Ethnic Fragmentation	Religious Polarization	Religious Fragmentation	Average Ethnolinguistic Fragmentation	Ethnic Dominance	Ethnic Inequality (GREG)	Ethnic Inequality (ETHNO)	Oil Gini
Panel regressions, 1940s-1980s										
log population	0.782	0.761	0.440	0.552	0.561	0.334	0.623	0.835	0.700	0.477
	(0.310)	(0.333)	(0.250)	(0.293)	(0.261)	(0.170)	(0.197)	(0.298)	(0.304)	(0.185)
log population × variable	0.149	-0.389	0.519	0.214	0.447	0.731	-0.320	-0.430	-0.143	0.124
	(0.150)	(0.516)	(0.381)	(0.308)	(0.506)	(0.316)	(0.183)	(0.385)	(0.493)	(0.204)
Observations	267	280	280	280	280	295	280	302	302	269
Number of clusters	53	57	57	57	57	60	57	62	62	54

Notes: 2SLS regressions with a full set of year and country fixed effects. Robust standard errors (clustered by country) are reported in parentheses. Unbalanced panels with one observation per decade. All interacted variables are demeaned before interacting to facilitate the interpretation of the main effect. In computing standard errors, Bangladesh, India and Pakistan are considered a single cluster. See the text and Appendix Table A-1 for definitions and details.

Table A-19: The Effect of Population on Conflict: 2SLS Estimates
Heterogeneous Effects II

	Dependent variable is fraction of decade in conflict according to Correlates of War								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Interactions of log population with...								
	Area	Average Agricultural Suitability			Cereal Suitability Index	Agriculture, Value Added (% of GDP)	Educational Attainment	Av. Constraints on Executive, 1940s	Av. Polity Score, 1940s
		(Climate)	(Soil)	(Soil × Climate)					
Panel regressions, 1940s-1980s									
log population	0.587	0.612	0.567	0.573	0.647	0.882	1.712	0.676	0.547
	(0.209)	(0.204)	(0.215)	(0.212)	(0.194)	(0.482)	(3.276)	(0.222)	(0.252)
log population × variable	-0.019	-0.162	0.113	-0.047	0.006	0.003	0.206	-0.011	-0.006
	(0.018)	(0.189)	(0.376)	(0.281)	(0.007)	(0.004)	(0.582)	(0.010)	(0.014)
Observations	302	295	295	295	302	168	303	278	278
Number of clusters	62	60	60	60	62	34	62	57	57

Notes: 2SLS regressions with a full set of year and country fixed effects. Robust standard errors (clustered by country) are reported in parentheses. Unbalanced panels with one observation per decade. All interacted variables are demeaned before interacting to facilitate the interpretation of the main effect. In computing standard errors, Bangladesh, India and Pakistan are considered a single cluster. See the text and Appendix Table A-1 for definitions and details.

Table A-20: The Effect of Population on Conflict: 2SLS Estimates
Sample Split I

	Dependent variable is fraction of decade in conflict according to Correlates of War									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Sample Split by...									
	GDP per capita	Ethnic Polarization	Ethnic Fragmentation	Religious Polarization	Religious Fragmentation	Average Ethnolinguistic Fragmentation	Ethnic Dominance	Ethnic Inequality (GREG)	Ethnic Inequality (ETHNO)	Oil Gini
Panel A: Panel regressions, 1940s-1980s & above median 1940 or ethnic dominance equal to one										
log population	0.174	0.565	0.888	2.278	2.278	0.777	0.247	1.027	1.171	0.555
	(0.416)	(0.334)	(0.428)	(2.253)	(2.253)	(0.335)	(0.143)	(0.562)	(0.591)	(0.314)
Observations	138	143	143	140	140	148	157	154	150	136
Number of clusters	28	30	29	28	28	30	33	31	30	27
Panel B: Panel regressions, 1940s-1980s & below median 1940 or ethnic dominance equal to zero										
log population	4.319	0.791	0.313	0.234	0.234	0.324	1.109	0.593	0.144	0.694
	(6.266)	(0.350)	(0.296)	(0.315)	(0.315)	(0.228)	(0.450)	(0.255)	(0.404)	(0.297)
Observations	169	137	137	140	140	147	123	148	152	171
Number of clusters	35	28	29	29	29	31	25	32	32	37

Notes: 2SLS regressions with a full set of year and country fixed effects. Robust standard errors (clustered by country) are reported in parentheses. Unbalanced panels with one observation per decade. Panel A presents panel regressions for countries where the variable in each column exceeds the median value in 1940 (or ethnic dominance is equal to one). Panel B presents panel regressions for the rest of the countries (below the median or ethnic dominance is equal to zero). In computing standard errors, Bangladesh, India and Pakistan are considered a single cluster. See the text and Appendix Table A-1 for definitions and details.

Table A-21: The Effect of Population on Conflict: 2SLS Estimates
Sample Split II

	Dependent variable is fraction of decade in conflict according to Correlates of War								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Sample Split by...								
	Area	Average Agricultural Suitability			Cereal Suitability Index	Agriculture, value added (% of GDP)	Educational attainment	Av. Constraints on Executive, 1940s	Av. Polity Score, 1940s
		(Climate)	(Soil)	(Soil × Climate)					
Panel A: Panel regressions, 1940s-1980s & above median 1940 or ethnic dominance equal to one									
log population	0.722	0.764	0.811	0.500	0.499	7.046	-0.393	0.356	0.648
	(0.460)	(0.418)	(0.270)	(0.265)	(0.144)	(15.280)	(0.508)	(0.260)	(0.382)
Observations	151	150	151	151	153	85	157	145	141
Number of clusters	31	31	30	30	31	16	32	30	29
Panel B: Panel regressions, 1940s-1980s & below median 1940 or ethnic dominance equal to zero									
log population	0.626	0.440	0.289	0.723	0.843	0.203	-4.958	0.852	0.731
	(0.257)	(0.151)	(0.284)	(0.369)	(0.386)	(0.160)	(8.530)	(0.243)	(0.320)
Observations	151	145	144	144	149	83	146	162	166
Number of clusters	32	30	31	31	32	18	30	34	35

Notes: 2SLS regressions with a full set of year and country fixed effects. Robust standard errors (clustered by country) are reported in parentheses. Unbalanced panels with one observation per decade. Panel A presents panel regressions for countries where the variable in each column exceeds the median value in 1940 (or ethnic dominance is equal to one). Panel B presents panel regressions for the rest of the countries (below the median or ethnic dominance is equal to zero). In computing standard errors, Bangladesh, India and Pakistan are considered a single cluster. See the text and Appendix Table A-1 for definitions and details.

Table A-22: Mexico: Calibrating the Population Impact of Decreased Malaria Mortality

	(1)	(2)
Panel A: Malaria mortality, suitability, and births		
Dependent variable is...	Births to population	Malaria death rate
	Just 1940 and 1980	1940
Predicted Mortality (only malaria)	-0.360	
	(0.342)	
Predicted mortality (Mexico)		2.535
		(1.866)
Sample	Cross-country	Mexico, State-level
Observations	95	32
R-squared	0.297	0.058
Panel B: Calibrated average increase in municipal population (in %)		
	Malaria mortality in 1940 declines to...	
Comorbidity assumption...	National average in 1960	State average in 1960
0.5 × Cross-country comorbidity	4.609	4.420
Cross-country comorbidity	9.410	9.029
1.5 × Cross-country comorbidity	14.41	13.83
	Size of the effect in baseline regression	
Estimated decline for average municipality relative to a municipality with zero malaria suitability	12.47	

Notes: Column 1 in Panel A is an OLS long-difference regression (just 1940 and 1980) with a full set of year and country fixed effects (equation (4.4) in the text). Column 2 is an OLS regression for Mexican states in 1940. The estimates in column 1 are used to calibrate the fertility response given the decline in malaria death rates. The estimates in column 2 are used to predict a municipal-level malaria death rate in 1940. Panel B presents the resulting cumulated population increase (in percent) at the municipal level when assuming a decrease in malaria mortality as indicated in each column, and with associated comorbidity effects as indicated in each row. Fertility effects are always included as predicted by column 1 of Panel A. In Panel B, cross-country comorbidity equals 1.96, from our estimates in Table A-16: the ratio of the coefficients for just malaria (column 7, -1.536) and the baseline predicted mortality (-0.782, column 1) in our alternative first-stage specifications for the effects on log population. When we suppose that municipality-malaria death rates decline to the state-level average (column 2 in Panel B), 227 municipalities in Oaxaca would have to exhibit an increase in mortality. We set the decline to zero in this case.

Table A-23: Mexico: Population and Protests per capita
2SLS Estimates
Heterogeneous Effects - Additional First Stages

Dependent variable is...	Non-harvest droughts × log population	Harvest droughts × log population	Non-harvest droughts × log population	Harvest droughts × log population
	(1)	(2)	(3)	(4)
Panel A: No controls				
Predicted mortality (Mexico)	5.765	1.627	2.076	0.675
	(0.416)	(0.195)	(0.403)	(0.146)
Predicted mortality (Mexico) × Non-harvest	-0.779	-0.020	-0.309	-0.010
	(0.042)	(0.012)	(0.048)	(0.013)
Predicted mortality (Mexico) × Harvest	-0.005	-0.626	-0.095	-0.323
	(0.096)	(0.051)	(0.094)	(0.063)
Observations	4,744	4,744	4,744	4,744
Panel B: All controls				
Predicted mortality (Mexico)	5.983	1.561	2.841	0.898
	(0.434)	(0.200)	(0.530)	(0.201)
Predicted mortality (Mexico) × Non-harvest	-0.738	-0.018	-0.315	-0.029
	(0.043)	(0.012)	(0.057)	(0.017)
Predicted mortality (Mexico) × Harvest	-0.063	-0.629	-0.207	-0.366
	(0.119)	(0.065)	(0.127)	(0.085)
Observations	3,566	3,566	3,566	3,566
Droughts x post-anti-malaria campaign dummy			✓	✓

Notes: Municipality-level OLS 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. 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 malaria suitability in 1940 and zero in 1960. Predicted mortality (Mexico) is equal to malaria suitability in 1940 and to 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. See the text and Appendix Table A-1 for definitions and details.

Table A-24: Mexico: Population, Shares of Population and Education
OLS & 2SLS Estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Share of pop. 20-39	Primary school	Literacy rate	Literacy rate (15-39)	Share of pop. 20-39	Primary school	Literacy rate	Literacy rate (15-39)
Panel A: Dependent variable is indicated in each column title (OLS)								
Predicted mortality (Mexico)	0.011	0.020	0.032	0.052	0.010	0.024	0.011	0.034
	(0.004)	(0.008)	(0.011)	(0.023)	(0.004)	(0.006)	(0.010)	(0.021)
Observations	4,318	4,222	4,325	4,227	3,496	3,443	3,540	3,450
Panel B: Dependent variable is violent protests per capita (2SLS)								
log population	10.189	11.089	12.561	12.873	12.622	13.049	13.810	13.936
	(3.615)	(3.132)	(3.039)	(3.025)	(4.368)	(4.130)	(4.196)	(4.177)
Control shown in each column title	7.186	22.902	6.838	5.507	4.188	22.467	3.934	3.014
	(7.232)	(6.286)	(1.450)	(1.263)	(7.074)	(7.709)	(1.655)	(1.409)
Observations	4,182	4,128	4,176	4,176	3,572	3,572	3,568	3,572
Controls x 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. Robust standard errors allowing for spatial correlation between municipalities within a radius of 35.9 km. In Panel B, log population is instrumented with predicted mortality (Mexico). Protests are counts of news stories about protests, expressed as a fraction of baseline population (per 100,000 people). Predicted mortality (Mexico) is malaria suitability in 1940 and zero in 1960. Predicted mortality (Mexico) is equal to malaria suitability in 1940 and to zero in 1960. Controls are distance to capital, distance to big cities, land quality index, log population in 1940, primary school in 1940, university enrollment in 1940, share of a municipality's area on a sedimentary basin, share of indigenous 1940 and historical conflict. See the text and Appendix Table A-1 for definitions and details.

Table A-25: Mexico: Population and Protests per capita
2SLS Estimates
Heterogeneous Effects - Robustness to Droughts Definition

	Dependent variable is violent protests per capita								
	Number of Months with Droughts					Decade Mean Precipitation / Long-run Mean Precipitation			
	1st percentile	2nd percentile	3rd percentile	4th percentile	5th percentile	All	Non-harvest	Harvest	Non-harvest & Harvest
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: No controls									
log population	11.367	11.403	10.955	11.117	10.771	12.349	12.657	12.588	13.815
	(3.754)	(3.683)	(3.623)	(3.602)	(3.623)	(3.268)	(3.333)	(3.380)	(3.714)
log pop × Non-harvest	0.521	0.503	0.492	0.456	0.446	35.925	40.035		46.747
	(0.196)	(0.183)	(0.171)	(0.157)	(0.150)	(11.098)	(10.966)		(13.262)
log pop × Harvest	-0.355	-0.294	-0.305	-0.247	-0.295			-1.920	-14.124
	(0.476)	(0.430)	(0.394)	(0.358)	(0.345)			(11.081)	(12.105)
Observations	4,744	4,744	4,744	4,744	4,744	4,744	4,744	4,744	4,744
Panel B: All controls									
log population	14.138	14.767	14.812	15.266	14.998	12.985	12.965	12.518	13.552
	(5.048)	(4.940)	(4.969)	(5.011)	(5.045)	(4.160)	(4.142)	(4.321)	(4.667)
log pop × Non-harvest	0.521	0.536	0.558	0.531	0.542	40.133	37.979		40.420
	(0.216)	(0.205)	(0.196)	(0.183)	(0.175)	(14.501)	(12.865)		(14.390)
log pop × Harvest	-0.250	-0.228	-0.219	-0.134	-0.165			5.080	-5.944
	(0.582)	(0.537)	(0.509)	(0.473)	(0.458)			(14.096)	(14.766)
Observations	3,566	3,566	3,566	3,566	3,566	3,566	3,566	3,566	3,566

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 expressed as a fraction of baseline population (per 100,000 people). Predicted mortality (Mexico) is malaria suitability in 1940 and zero in 1960. Predicted mortality (Mexico) is equal to malaria suitability in 1940 and to 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. The ratio of decade mean precipitation and the long-run mean precipitation is censored at one. Controls are distance to capital, distance to big cities, land quality index, log population in 1940, primary school in 1940, university enrollment in 1940, share of a municipality's area on a sedimentary basin, share of indigenous 1940 and historical conflict. See the text and Appendix Table A-1 for definitions and details.