INSTITUTE for REPLICATION

No. 35
I4R DISCUSSION PAPER SERIES

# A Comment on Dincecco et al. (2022): Pre-Colonial Warfare and Long-Run Development in India

Rachel Forshaw

Tim Ölkers

Ritika Sethi

Manali Sovani

June 2023

I4R DISCUSSION PAPER SERIES

I4R DP No. 35

# A Comment on Dincecco et al. (2022): Pre-Colonial Warfare and Long-Run Development in India

**Rachel Forshaw[1], Tim Ölkers[2], Ritika Sethi[3], Manali Sovani[4]**

[1]*Heriot-Watt University, Edinburgh/Scotland*
[2]*University of Göttingen/Germany*
[3]*Rice University, Houston, TX/ USA*
[4]*Tufts University, Medford, MA/USA*

JUNE 2023



ISSN: 2752-1931

# A comment on Dincecco et al. (2022): Pre-colonial warfare and long-run development in India *

Rachel Forshaw† Tim Ölkers Ritika Sethi Manali Sovani

April 26, 2023

**Abstract**

We test the reproducibility and replicability of Dincecco et al. (2022), which reports a positive relationship between pre-colonial interstate warfare and long-run development patterns across India. Overall, we confirm that all of the study's estimates are computationally reproducible by using both the provided replication package in Stata and code written by the present authors in R. We test for and find no evidence of data manipulation in the final datasets. Concerning direct replicability, we consider different ways of measuring distance to conflicts and also alternative proxies for both the dependent variable and variables which capture channels by which the main effects operate. We are able to replicate the magnitude and significance of the estimated coefficient on conflict exposure in most of the tests, noting that while most estimates are substantively in line with the original study, some alternative measures of distance to conflict imply different magnitudes for estimates, and proxy estimates are sensitive to both the time period and type of conflict considered.

KEYWORDS: institutions, long-run development, path dependence, public goods

JEL CODES: H11 - Structure, Scope, and Performance of Government, N45 - Asia including Middle East, O11 - Macroeconomic Analyses of Economic Development, P48 - Political Economy; Legal Institutions; Property Rights; Natural Resources; Energy; Environment; Regional Studies

*Authors: Forshaw (†corresponding author): Heriot-Watt University. Email: r.forshaw@hw.ac.uk. Ölkers: University of Göttingen. Email: tim.oelkers@uni-goettingen.de. Sethi: Rice University. Email: ritika.sethi@rice.edu. Sovani: Tufts University. manali.sovani@tufts.edu. This replication was part of the Calgary Replication Games organised by the Institute for Replication (https://i4replication.org/). We thank Abel Brodeur and participants of the Calgary Replication Games, and anonymous referees for their suggestions.

# 1 Introduction

Dincecco et al. (2022), henceforth DFMM, investigate the relationship between pre-colonial interstate warfare and long-run development patterns across India. They construct a new geocoded database of historical conflicts on the Indian subcontinent and find a robust positive relationship between pre-colonial land battle conflict exposure and economic development. In their preferred specification, the authors find that pre-colonial conflict exposure to land battles within 250km of the district centroid between the years 1000 and 1757 is associated with increased contemporary economic development as measured by district-level luminosity averaged between 1992 and 2010. They argue that districts more exposed to pre-colonial conflict experienced greater early state-making which increased the powers of local government institutions. The greater power a local government institution held, the more the promotion of local long-term economic development through the provision of domestic security and investments in physical and human capital. In the long run, the authors argue, this led to higher levels of development and less political violence.

The goal of our study is to replicate the all results of DFMM, to add further extensions as well as robustness checks to the study. We define a positive replication as an estimate of the same sign (positive/negative) and significance (significant/not significantly different from zero) as that reported the original paper. This definition of course precludes difference in the magnitudes of estimates, which we discuss in the text. Our study first successfully replicates 100% of the main findings of the authors directly using code in Stata and data provided by the original study. We also confirm that 100% of findings can be replicated in an alternative software, R. Given that raw data is not provided by the authors, we calculate the distributions of first digits in the prepared data provided by the study and compare them with the distributions we would expect from non-manipulated data. Using this technique, we find no evidence of manipulation of the data in 100% of tests.

To assess direct replicability, we use DFMM's methodology on alternative data. Firstly, we consider alternative proxies for conflict exposure using data that are provided by the authors, but not used in the original paper. We find that some estimated coefficients on conflict exposure are sensitive to the type of conflict (58% of results replicated) and time period over which the conflict exposure is considered (73% of results replicated). We also explore alternative ways of measuring conflict exposure using data from the Historical Conflict Event Dataset (Miller and Bakar 2023). We are able to replicate the sign and significance of estimates in 100% of alternative measures of conflict exposure, although we note that magnitudes differ. We also find that the number of conflicts varies greatly with time, with most recorded conflicts being registered in period 1500-1757, the period of the Mughal empire. We examine the extent to which time heterogeneity in recorded conflicts translates to time heterogeneity in DFMM's main results and channels by which their main effects operate.

Finally, we re-examine one of the channels by which DFMM's main effects operate, the relationship between pre-colonial conflict exposure and contemporaneous political violence levels. To do so, we use an alternative proxy for political violence, including data provided by the original study and new data from Sundberg and Melander (2013). We find that the results for this channel are sensitive to the choice of proxy and time period considered in 50% of tests.

The remainder of the paper is organised as follows: section 2 discusses the results of tests of computational reproducibility. Section 3 covers tests of direct replicability and section 4 concludes.

## 2 Computational reproducibility

This section considers the computational reproducibility, or the ability to duplicate the results of a study using the same data and procedures as were used by the original investigators. We first show reproducibility of the main results using both the

Stata code provided by DFMM and recoded by the present authors in R in section 2.1. Controlling the reproducibility with another program is of interest because Sauerbrei et al. (2006) found that for certain regression techniques differences to Stata and R programs are noted. Secondly we test for data manipulation in the datasets provided in DFMM, since only final datasets are available, and find no evidence of data manipulation.

## 2.1 Stata and R reproducibilty

Both the code and full datasets are provided by the authors and published on the Economic Journal website, available here. We reproduce the paper's main result, an ordinary least squares regression of luminosity on pre-colonial conflict exposure, and main robustness check, which uses two stage least squares and instruments for pre-colonial conflict exposure.

For this analysis, we rely on the same specifications as DFMM, the OLS specification is:

$$Y_{i,j} = \beta \text{ConflictExposure}_{i,j} + \lambda \text{PopDensity}_{i,j} + \mu_j + \mathbf{X}'_{i,j}\phi + \epsilon_{i,j} \tag{1}$$

where $i$ indexes districts in equation 1 and $j$ indexes states in modern-day India. $Y_{i,j}$ measures local economic development in terms of luminosity, $\ln(0.01 + \text{Luminosity}_{i,j})$. $\text{ConflictExposure}_{i,j}$ measures pre-colonial conflict exposure, the variable of interest. $\text{PopDensity}_{i,j}$ controls for log population density, $\mu_j$ are state fixed effects and $\mathbf{X}'_{i,j}$ is a vector of controls for geographic features including latitude, longitude, altitude, ruggedness, precipitation, land quality, dry rice suitability, wet rice suitability, wheat suitability and malaria risk.

Table 1 shows the paper's main results, reproduced by using the authors' Stata code ('Original Study') and reproduced in the coding language R (R Core Team (2021)) using command 'lm' for specification (1) and 'felm' command from package 'lfe' (Gaure (2022)) for the specifications containing state fixed effects (2 and

3). Table 2 reproduces the paper's main robustness check, which uses a district's proximity to the Khyber pass as instrument for pre-colonial conflict exposure, since the Khyber pass was the main route for invaders coming from Central Asia to India. The R reproduction uses command 'ivreg' whereas the Stata code uses 'ivreg2' (Baum et al. (2002)). In both reproductions, we found no discrepancies between the original paper and either the Stata or R reproductions in terms of point estimates. Small differences arise only in the handling of standard errors, since Stata adjusts standard errors for small samples. However, using White HC1 standard errors using the 'lmtest' package Zeileis and Hothorn (2002) in R results in very similar standard errors and does not alter the significance of the estimates.

## 2.2 Distribution of first digits

DFMM provide the final datasets, but not the base datasets nor the code to create them. As such, it was not possible to check data definitions or to recode key variables. Instead, we write a Python (Python Core Team (2015)) routine that extracts the leading digits (excluding zeros) in each dataset and compares their relative frequency distribution against Benford's law, which describes the expected relative frequency distribution for leading digits of numbers in datasets. Comparing data to this theoretical distribution is a technique used to look for fraud and manipulation in financial records and other datasets in forensic accounting. This test is most likely to capture data fraud under the condition that observations have been added, edited, or removed in a way that does not conform to the Benford distribution; see Durtschi et al. (2004) for a discussion. Figure 1 shows the expected distribution in non-fraudulent data. We calculate distributions for all non-assigned variables in all 19 datasets provided by DFMM and compare these distributions to the expected distribution. Figure 2 shows examples of calculated distributions compared with the expected distributions for the variables used in the main specifications of DFMM. Following Azevedo CDS (2021) we define a calculated distribution as not conforming to the expected distribution when the mean squared error (MSE) is in excess of 0.015.[1] Panels A-C, which show measures of Luminosity, Conflict Exposure and Population Density, all exhibit MSEs less than the critical value and with distributions that clearly align to the expected frequencies. Panel D shows an example of data in which the calculated frequencies of first digits do not align to the Benford distribution. Since this is latitude data we would not expect it to do so, given that the Indian subcontinent can only lie within given latitudes, i.e., this data is assigned. Repeating this process for all non-assigned, numeric variables in the 19 datasets provided by DFMM, we find no evidence of data manipulation.

# 3 Direct replicability

In this section, we test the ability to duplicate the results of DFMM using new data but the same procedures as were used by the original investigators. All direct replicability tests focus on the measurement of and types of proxies used for conflict exposure. Section 3.1 reports estimates using alternative proxies provided in the DFMM replication package but not reported by authors in either the main text or online appendix. In section 3.2 we explore alternative ways to measure proximity to conflict exposure using data from the Historical Conflict Event Dataset. Section 3.3 considers estimates when different time periods of the conflict exposure proxy are used with data provided by DFMM but not reported. In section 3.3, we also consider the replicability of channels analysis by DFMM, which considers the relationship implied by their theoretical framework between pre-colonial conflict exposure and eventual political violence levels. Finally in section 3.4 we consider a proxy of political violence not considered in the original article, using data from the Uppsala Conflict Data Program (UCDP) of Sundberg and Melander (2013).

[1]Other measures that are commonly used for anomaly detection include Chi-squared and Z-tests which we do not consider since all datasets are large, and pass MSE and visual inspections. For the same reason we also do not test final digits, a lesser-used test for data fraud.

### 3.1 Alternative proxies for conflict exposure

DFMM utilise land battles between 1000 and 1757 within 250km of the district centroid as the main proxy for conflict exposure. The authors use historical conflict from Jaques and Showalter (2006), which they geocode to a create measure of exposure to individual conflicts for Indian districts using equation 2. As a robustness check, we use alternative proxies of conflict exposure included in the dataset provided by DFMM. These measures are All conflicts, Multi-day, Multi-year, Naval, One-Day, Sacking & Razing, Seige and Storming battles. Table 3 reports OLS the estimates from the original study (row 1) and the estimates using alternative proxies (rows 2-9) and table 4 the second stage IV estimates. Of the alternative proxies, we replicate the sign and significance of the original paper's results in 60% of tests. However, we note that only the One-Day battles and All conflicts proxies replicate the results of DFMM closely since the magnitudes of estimates are much larger in specification (1) for Multi-year, Naval and Sacking & Razing. A number of the coefficient estimates using alternative proxies are not significantly different from zero in the both specifications once state fixed effects and geographic controls are included (columns 2 and 3, respectively), encompassing Multi-day, Multi-year, Naval (of which the estimates are negative in the IV estimates), Seige in the OLS specification and Storming in both.

### 3.2 Alternative measures of conflict exposure

DFMM define the exposure to conflict as the sum of the inverse distance between each district centroid and pre-colonial conflicts. As robustness checks, the authors further use alternative radii cutoffs to define conflict exposure (table A.15, Online Appendix). They also explore a variable end-date cut-off which allows them to include exposure to conflicts after 1757 but before British conquest of a district (table A.16, Online Appendix). In both cases, they find that the coefficient estimates are very similar in magnitude and significance to the main estimates across both checks.

This section explores further robustness checks by using use alternative measures of exposure to conflict.[2] We begin by replicating the measure used in the paper, and then consider several alternative measures.

DFMM define the exposure to conflict as the sum of the inverse distance between each district centroid and pre-colonial conflicts that occurred between 1000 and 1757 within a radius of 250 km:

$$\sum_{c \in C} (1 + \text{ distance}_{i,c})^{-1} \tag{2}$$

where $\text{distance}_{i,c}$ is measured from the centroid of district $i$ to the location of conflict $c$. We use data from the Historical Conflict Event Dataset (HCED) of Miller and Bakar (2023) and construct the measure of exposure to conflict in equation 2, keeping only land battles that occurred in the Indian subcontinent[3] between 1000 and 1757, and within 250 km. We calculate the geographical distance for each district and conflict pair. We do so by measuring the length of the shortest path between the two points along the surface the earth, as approximated by the method of Vincenty (1975) to calculate distances on a reference ellipsoid. Figure 3 compares distributions of the original and reconstructed measures using histograms. They are not identical but close in distribution, with a notably fatter right tail in the replication. Table 7 compares the number of HCED conflicts with the number of conflicts calculated using data provided by DFMM.[4] There are fewer conflicts in total in the HCED data, concentrated between 1000-1100 and post 1500.

In Table 5, we compare how the results of the main specification in (1) differ when we use the reconstructed measure (panel B) in place of that provided by DFMM (panel A). The main parameter of interest, i.e., the effect of exposure to

[2]We thank an anonymous referee for this suggestion.

[3]Following Dincecco et al. (2022) this includes conflicts that took place in modern day India, plus the border nations of Bangladesh, Bhutan, Mayanmar, Nepal, Pakistan and Sri Lanka.

[4]DFMM provide data on the latitudes and longitudes of both the district centroids and the start year of conflicts. Files prefixed *IND_adm2* contain the district centroids and file *formapping.csv* contain the latitudes, longitudes and start dates of conflicts. We verify in ArcGIS that the centroids of districts are correctly measured.

conflict on present-day development, is still positive and statistically significant but at a marginally smaller magnitude. We find similar minor reductions in the magnitude of the effect of exposure to conflict on present-day development when we include state fixed effects and geographical controls. Similarly, we check how the results of the IV analyses may change when we use the reconstructed measure. Table 6, panel B, presents the second stage estimates. The parameters of interest have the same sign and statistical significance but lower magnitudes. As discussed in the preceding paragraph, measures using HCED data do not exactly match those of Dincecco et al. (2022). DFMM also do not clarify the exact algorithm that they rely on to measure distance. To understand whether the different data or different distance measurements are driving the change in magnitudes, we repeat the analysis using data from DFMM and report the OLS estimates in panel B′ of table 5, and 2SLS estimates in panel B′ of table 6. Since DFMM only provide a crosswalk for All Conflicts, the relevant comparison is row 2 of tables 3 and 4. Here we see that significance, signs, and magnitudes are almost exactly equal. We conclude that it is therefore most likely that the differences in the data that are driving the slight differences in estimated magnitudes.

We now summarize the alternative measures of distance to conflict that we explore. To reduce the measure's sensitivity to any single conflict, DFMM add one to $\text{distance}_{i,c}$ before taking the inverse. They argue that excluding this would mean that a district in which a conflict took place very near to the centroid would receive a large conflict exposure value, regardless of its proximity to any other conflicts. We explore how the results may change if we omit adding one to $\text{distance}_{i,c}$, i.e.,

$$\sum_{c \in C} (\text{distance}_{i,c})^{-1} . \tag{3}$$

Here, we still follow their benchmark of including conflicts that occurred between 1000 and 1757, and within 250km of the district. Table 5 reports that the effect of exposure to conflict is still statistically significant but of marginally lower magnitude. Table 6 compares the results of the IV analyses. The magnitudes of the parameter of interest are close in magnitude between using the original and alternative measures and of the same sign and siginificance.

In the second and third alternative measures, we use the baseline measure as in (2) but explore using different distance units: miles and 100km. For example, if originally a distance from conflict $a$ to district centroid $b$ measured $x$ kilometers, we now measure $0.62x$ miles and $x/100$ kilometers, respectively. Comparing OLS and IV analyses using these measures, panels D and E of tables 5 and 6 show that estimates are smaller in magnitude in both the OLS and IV specifications, respectively. This exercise illustrates that the magnitude of the effect is sensitive to the unit measurement, as would be expected, but given that equation 2 is non-linear the interpretation of magnitudes is not straightforward. However, the sign and significance of the estimates are unaffected.

Fourth, we explore a Gaussian transformation of $distance_{i,c}$ (measured in 100 kilometers):

$$\sum_{c \in C} \exp\left(-distance_{i,c}\right) \tag{4}$$

which captures distance decay, a measure used in geography to describe the decline of influence on cultural or spatial interactions between places as distance increases Pun-Cheng (2016). This measure of conflict exposure is positively correlated with the original exposure measure of DFMM. Panel F of tables 5 and 6 show that the sign and significance of the estimated coefficients are unaffected, but the magnitude is significantly lower.

Finally, we define a simple count-based measure of exposure to conflict, which is calculated as the number of conflicts within 250km of the district centroid. Panel G of tables 5 and 6 show that the sign and significance of the estimated coefficients are again unaffected, but the magnitude is significantly lower. This measure of conflict exposure is appealing as it has a straightforward interpretation: an increase of one conflict within the proximity of the district increases contemporaneous economic activity (as proxied by luminosity) by between 2 and 4 percentage points, an economically significant finding.

In conclusion, 100% of tests using alternative measures of conflict exposure replicate the sign and significance of the original paper's results. However, all estimates have lower magnitudes of the effect of exposure to conflict on present-day development than the original measure, although the interpretation of differences is complicated by the non-linearity of the distance measure. Using a simple count of conflicts within 250km of the district centroid confirms that the estimated effect is positive and economically significant.

### 3.3 Different time periods for pre-colonial conflict exposure

The replication of section 3.2 highlighted the heterogeneity in the number of conflicts recorded per century in DFMM's data. In this section, we therefore test the sensitivity of DFMM's results to the period of time over which pre-colonial conflict exposure is measured. In tables 9 and 10 we replicate the main OLS results (table 1 of DFMM) and IV specification (table 2 of DFMM), respectively, but break down the pre-colonial conflict exposure variable into 100 year periods. Concretely, columns 1 are for exposure during years 1000-1757 (ie, replicating the original study). Columns 2 reports the estimates for the time period 1000-1100, columns 3 for exposure during years 1101-1200 and so on until column 8 for years 1601-1700. Note that for the time period 1400-1500, estimates are omitted as there were no recorded conflicts in the DFMM data. We replicate the OLS results in terms of sign and binary significance in 100% of tests, and 83% of the IV specification tests. However, we see great heterogeneity in estimates over 100 year time blocks. In the OLS results of table 9, in later time blocks, 1501-1600 and 1601-1700, we find estimates of a much smaller magnitude. Earlier time periods, 1000-1100 to 1301-1400 have estimated coefficients that are larger. We see a similar pattern in the IV results in table 10, and in this specification, time block 1601-1700 now shows an insignificant effect. We note that DFMM restrict the time period for the conflict data to the sub-period of 1500 to 1757 and report their main findings (i.e., replicating tables 1 and tables 2) in their online appendix (their tables A.13 and A.14) and report robust results.

Table 7 in our paper reports the number of pre-colonial conflicts while Table 8 reports the means of conflict exposure broken down by 100-year time periods. Both the number of conflicts and average exposure is much higher after 1500. This could mean that the observed pattern of coefficients could be due to a size effect owing to the lower number of conflicts before 1500, perhaps due to selection in the recording of conflicts. It could also be that conflicts before 1500 had a larger effect on current economic conditions due to decreasing returns to conflict exposure. Either mechanism would rationalize smaller estimates, and this analysis is unable to distinguish between the two.

In order to analyse this further, we are able to examine the time heterogeneity in channels by which their main effect operates according to DFMM's theoretical framework. This sub-period of time is pertinent for another reason: since the Mughal empire lasted between 1500 to 1757, this was the duration when India was at its wealthiest and it was a time period where Mughal emperors fought multiple wars either to unite India under Mughal rule (like Akbar) or to keep enemies at bay (Like Aurangzeb against the Marathas). We therefore utilise this sub-period measure to estimate the remaining results of the paper, i.e. their tables 3-9.

Firstly, table 11 replicates table 3 in the original study, which estimates the relationship between pre-colonial state-making and pre-colonial conflict exposure. An important prediction of the theoretical framework of DFMM is a positive relationship between the two as measured by the number of important Mughal sites, and districts incorporated into the Mughal empire by rulers Babur and Akbar. We replicate the original study in the upper panel of table 11, and in the lower panel we use the later time period for pre-colonial conflict (years 1500-1757). We find a significant and positive relationship for important Mughal sites, and districts incorporated into the Mughal empire by rulers Babur and Aurangzeb. We also note that the estimated magnitude of the effect is greater during the later years, 1500-1757, than for earlier years.

Table 12 replicates the regression of colonial fiscal development on pre-colonial conflict exposure. DFMM report a positive and significant relationship between fiscal development and pre-colonial conflict exposure between 1000-1757, which the authors argue is suggestive evidence that pre-colonial conflict exposure played a role in colonial-era state-making. We reproduce their results in the first row of table 12. Using the later time period for conflict exposure measure provided in the original study, 1500-1700, in row two we show results that are broadly consistent with those of the original study, which shows similar patterns across the authors' different measures. Such measures include different scaling of the available tax revenue in 1881, by area and persons across states with direct rule (British India) or indirect rule (Princely states) and tax revenue in 1931 scaled by area and by person. Again, we find that estimated magnitudes are marginally greater during the later years, 1500-1757.

Table 13 replicates table 5 of Dincecco et al. (2022), which examines the relationship between pre-colonial conflict, colonial and post-colonial conflict. Regressing local exposure to colonial and post-colonial conflicts on pre-colonial conflict exposure, the authors find a positive and significant relationship with colonial conflict exposure between 1758-1839, a negative relationship with post-colonial conflict exposure and no relationship with colonial conflict exposure during 1840-1946, all respectively measured by land battles and all conflicts proxies. We find these results are robust to the use of the later time period for pre-colonial conflict exposure and that estimated magnitudes are greater during the later years, 1500-1757.

Row one of table 14 reports the replication of table 6 of DFMM, which estimates a negative relationship between pre-colonial conflict and post-colonial political violence. A possible explanation for this might be that regions that had experienced higher conflict under the Mughals and had been united under one ruler were easier to govern by the British and the same political and geographic stability was inherited by the newly independent India after 1947. Row two shows the replication with the subset of conflicts, finding similar results. While the relationship with linguistic fractionalisation (column 2) is no longer significant, the relationship with political violence remains robust. Note that for column 2, which measures local Maoist control in 2003, the number of observations in the replication is lower than the original study (293 versus 395) since this dataset does not include the later time period data for pre-colonial conflict, and we are only able to match observations for a subset of the original data.[5] Nevertheless, we estimate a relationship that is similar to the original study.

Table 15 reports the results from replicating DFMM's table 7, estimating the relationship between pre-colonial conflict and irrigation infrastructure. They find a large positive relationship which they argue is consistent with their theoretical framework which predicts greater state-making for areas with more conflict exposure, resulting in more investment in physical capital. The relationship with the share of non-agricultural workers in 2011 (% Non-agriculture, column 4) remains robust to the later time period for pre-colonial conflict exposure and is higher in magnitude for the later time period. However, column 1 shows that the positive relationship with the proportion of area sown with canal irrigation in 1931 (% irrigated) is no longer significantly different from zero when using the later time period for conflict exposure. In columns 2 and 3 we are not able to directly replicate DFMM's results as the data containing irrigation rates and crop yields does not contain the conflict exposure data for 1500-1757. We are only able to match data for 208 of the original 271 observations. With this subset of data, the relationship with irrigation rates averaged between 1956-87 (column 2), and the relationship with crop yield (column 3) is no longer significantly different from zero. However,

[5] It is likely that it would be possible to recover all observations were a state and district crosswalk provided in the replication package.

it impossible to ascertain whether this is due to the later time period considered for conflict exposure, or due to missing observations.

Table 16 replicates table 8 in DFMM, which estimates the relationship between pre-colonial conflict exposure and district-level literacy rates under British colonial rule. DFMM estimate no relationship for literacy rates in 1881 and 1921 but strong positive relationships in 1961-91 and 2011 (upper panel of table 16). We are able to directly replicate the results in columns 1 and 2 using the conflict exposure data from 1500-1757 provided by the authors, and also find no significant relationships. For literacy rates in 1961-91 and 2011, the authors do not provide sufficient data in the replication package to create these estimates, as the data containing literacy rates does not include conflict exposure data for the limited time period. We create this variable but are unable to recover as many observations as the original study for columns 3 (we recover 264 observations versus 271 in the original study) and 4 (541 recovered versus 626 in the original study). Using this subset of data, we no longer find a significant relationship between literacy rates and pre-colonial conflict exposure, however we are unable to disentangle whether this is due to the effect of dropped observations or the conflict exposure period used.

Tables 11 to Table 17 replicate the results of DFMM for a different time period, 1500-1757, the time of the Mughal empire. Overall, we find estimates of the same sign and significance during this time period in 92% of replications. We also find estimated magnitudes for estimates of channels are in general higher for this time period, providing suggestive evidence that the heterogenous time effects estimated in table 9 and 10 are more likely due to size effects than diminishing returns to conflict exposure.

## 3.4 Alternative political violence data source

As part of their paper, DFMM investigate potential pathways via which their estimated effects work. One of such investigations tests the relationship between pre-colonial conflict exposure and eventual political violence levels. To this end, the authors use data provided by the Armed Conflict Location & Event Data (ACLED) Project (Raleigh et al. 2010) to measure political violence today as outcome variables in their Table 6, column 1. The authors use two measures: the number of fatalities per district between 2015 and 2018 and local Maoist control in 2003 on pre-colonial conflict exposure. They find results consistent with a prediction of their theoretical framework, that pre-colonial conflict exposure should be a negative and significant relationship between pre-colonial conflict exposure and (eventual) political violence levels. We replicate their estimates in column 1 of table 18.

To assess the robustness of these estimates, we use conflict data provided by the Uppsala Conflict Data Program (UCDP) (Sundberg and Melander 2013) to re-estimate the effect of organised violence per district between 2015 and 2018 on pre-colonial conflict exposure (see equation (1)). The UCDP data covers individual events of organised violence, which they define as the phenomena of lethal violence occurring at a given time and place. In other words, we use a different dataset that includes and defines incidents of conflict differently as a robustness check. The UCDP events are sufficiently fine-grained to be geo-coded down to the level of individual villages, with temporal durations disaggregated to single, individual days. This is a similarity with the ACLED data, which is very detailed in terms of geographic coverage. However, micro-level studies find that analyzed based on conflict zones, like that of Dincecco et al. (2022), can misconstrue the correlates and patterns of internal conflict (Kalyvas (2008)). Hence, using a different conflict data source that also has a different measure of conflicts and re-estimating the results of DFMM on the district level helps to assess the robustness of DFMM results. Furthermore, we can investigate the sensitivity of the new conflict dataset by changing the time periods of our new conflict measure.

We use the same time period in the UCDP as Dincecco et al. (2022) use with their ACLED data source, 2015 to 2018. During this time period, there were incidents in 36 states and 659 districts, according to ACLED data used by DFMM. The UCDP data includes fewer incidents in 28 states and 288 districts during the same time period.

We present the results using this different conflict data source in column 2 and 3 of table 18. Using the UCDP data, we no longer find a negative and significant relationship between pre-colonial conflict exposure and contemporary conflicts that happened using the same time period (between 2015 and 2018) and geographical level (district) as the DFMM paper. If we include all conflicts between 2001 and 2021 (column 3, table 18), we do confirm the negative and significant relationship between pre-colonial conflict exposure and contemporary conflicts but note that it is sensitive to changes in time period considered.

# 4 Conclusion

Table 19 consolidates the results of all tests discussed in this paper. We confirm direct reproducibility of 100% of the main results using both the provided replication package in Stata and a routine that we write in R (table 19, Computational Reproduction). By testing the distribution of first digits and comparing it to an expected distribution, we find no evidence of data manipulation in any datasets provided by the authors (First Digits). For direct replicability, we consider alternative measures of conflict exposure using the Historical Conflict Event Dataset (Miller and Bakar 2023). As DFMM do not include code and intermediate data for their conflict exposure measure, we are unable to replicate an exact match for the measure of conflict exposure. However, we replicate the sign and significance of the original findings in 100% of tests. While the magnitudes of estimates differ for alternative measures, we confirm a significant positive association between long-run economic development and pre-colonial conflict exposure. We also examine alternative proxies for conflict exposure provided by the authors. DFMM argue that such alternative proxies were more likely to capture battles that affected the capital stock, diminishing the proposed mechanisms. Nevertheless, in 60% of tests (*Total, alternative conflict proxies*) we are able to replicate the sign and significance reported by DFMM, but note that some magnitudes of estimates also differ from the original paper. When considering different time periods between years 1000-1757 for the conflict exposure proxy, we are able to replicate results in 92% of tests (*Total, different time periods (main results)*, but report heterogeneity over in the magnitude of estimated coefficients along 100 year time blocks with larger estimates concentrated in the pre-1500 time period. Analysis of the number of conflicts and mean of conflict exposure shows that there is some evidence for a size effect given that many more conflicts are recorded in the post-1500 period, however we cannot rule out diminishing returns to conflict exposure in later periods. To examine this latter explanation, we replicate all results focusing on the channels by which their main results operate for the 1500-1757 period. In this analysis, we replicate their results in 67% of tests (*Total, different time periods (mechanisms results)*. For pre-colonial era state-making, colonial fiscal development, post-colonial conflict and post-colonial violence we find relationships that are broadly aligned with those estimated by DFMM. For irrigation infrastructure, literacy, presence of high schools and infant mortality, we do not find significant results using a later time period for conflict exposure. However, we note that a number of these results may be replicable if the authors provided a crosswalk for state and district between all datasets in their replication package. In general, we find higher magnitudes for estimates in the post-1500 period, suggestive evidence that size effects may play more of a role in time heterogeneity in the effects of conflict exposure on economic development than diminishing returns to exposure. Using an alternative proxy for political violence from the Uppsala Conflict Data Program (UCDP) (see Sundberg and Melander (2013)), we are not able to replicate estimates using the time period for the dependent variable used in DFMM, but are able to replicate the findings using a larger window including more recent data. Taken together, we confirm direct replicability in 70% of direct tests (Total Direct). Taking all results together (*Grand Total*), we find that 76% of tests have a positive replication result.

Contributors to Economic Journal are required to provide all the components necessary for others to duplicate the results of a study using the same materials and procedures as were used by the original investigator. Having this comprehensive criteria of the journal in mind, and based on the replication package provided by the authors, we argue that ease of replicability could be increased by the inclusion of a crosswalk linking observations in main and auxiliary datasets used in the channels analysis, and the provision of code and intermediate data for the construction of the conflict exposure measure. However, we conclude that the results of Dincecco et al. (2022) are replicable and the replicated estimates are substantively in line with the original study.

# 5 Figures

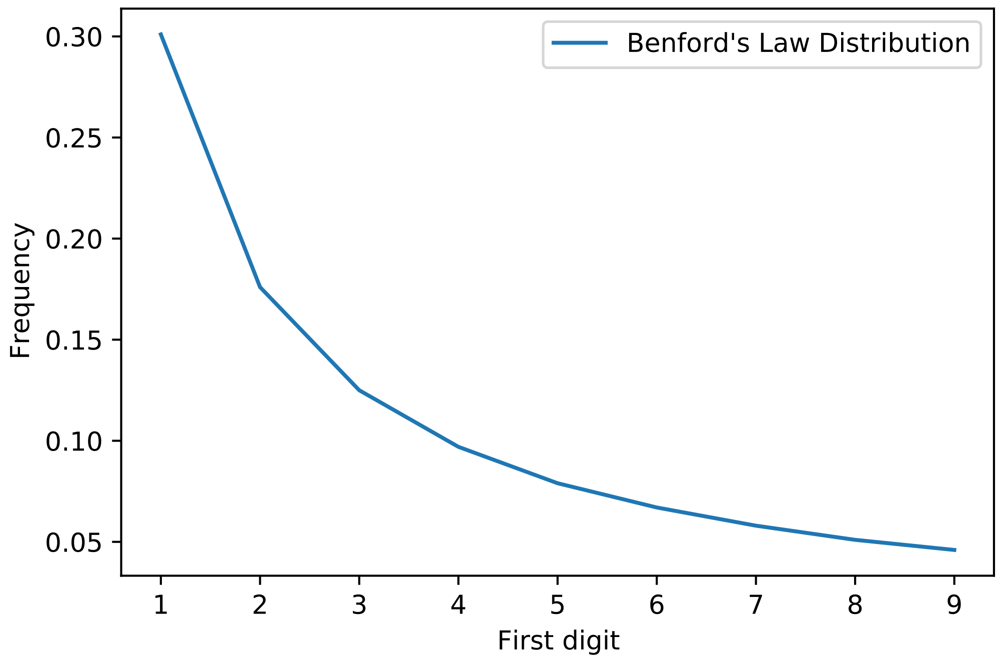
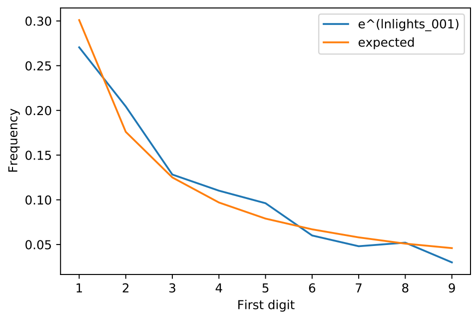
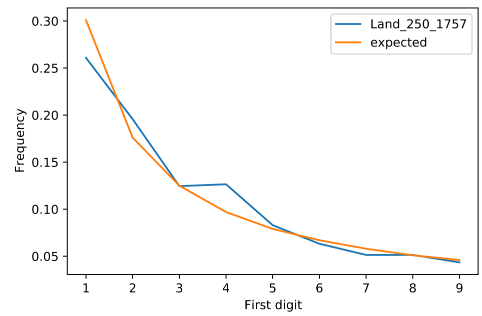



Figure 1: Benford's Law Distribution

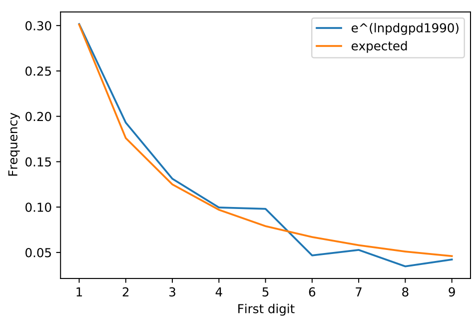
Expected distribution of all leading digits according to Benford's Law, which we find for all non-assigned variables in all 19 datasets provided by DFMM.



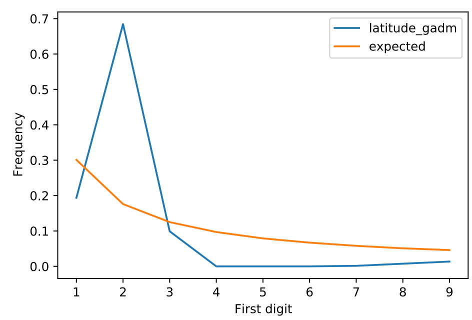

A: Luminosity




B: Conflict Exposure



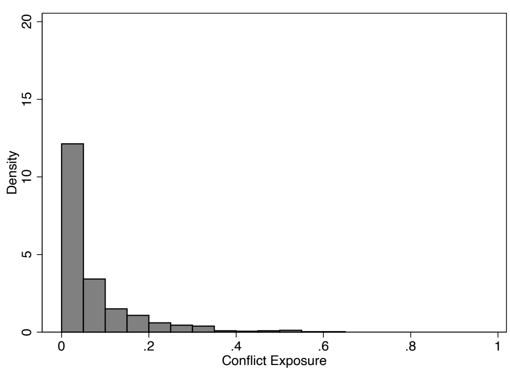

C: Population Density



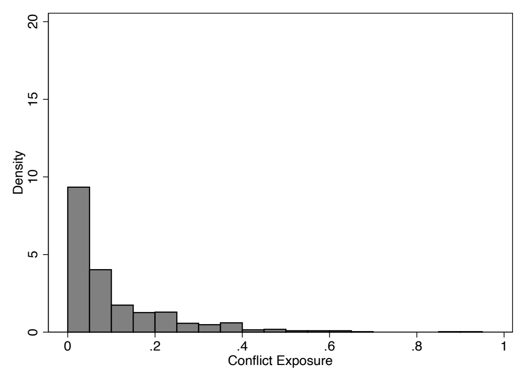

D: Latitude

Figure 2: Empirical Benford Distributions

Panel A: Proxy for economic development as measured by luminosity, the dependent variable, exponentiated to recover the distribution of digits i.e., $e^{\ln(1+Luminosity_{i,j})}$ in district $i$ of state $j$. This data fits the expected distribution with a mean squared error (MSE) of 0.0003. Panel B: Conflict exposure, the main variable of interest, as measured by Land Battles between 1000 and 1757 within a radius of 250km. This data fits the expected distribution with a mean squared error (MSE) of 0.0003. Panel C: Population density, control variable, exponentiated to recover the distribution of digits i.e., $e^{\ln(PopDensity_{i,j})}$. This data fits the expected distribution with a mean squared error (MSE) of 0.0002. Panel D: Latitude as measured using district centroids. This data fits the expected distribution with a mean squared error (MSE) of 0.03, which fails the Benfold test, which we would expect given the assigned nature of the data.




A: Conflict Exposure Measure (original)




B: Conflict Exposure Measure (replication)

Figure 3: Conflict Exposure Measure Distributions

Panel A (left): distribution of the Conflict Exposure measure by district provided by Dincecco et al. (2022). Panel B (right): Replication using data from the Historical Conflict Event Dataset (Miller and Bakar 2023).

# 6 Tables

Table 1: *Pre-Colonial Conflict and Economic Development: Main Results*

| | Original Study | | | R reproduction | | |
|---|---|---|---|---|---|---|
| Dependent Variable | ln(0.01 + Luminosity) | | | | | |
| | (1) | (2) | (3) | (1′) | (2′) | (3′) |
| Pre-colonial conflict exposure | 3.713*** | 1.601*** | 1.465*** | 3.713*** | 1.601*** | 1.465*** |
| | (0.305) | (0.380) | (0.370) | (0.305) | (0.380) | (0.370) |
| Population density | Yes | Yes | Yes | Yes | Yes | Yes |
| State FE | No | Yes | Yes | No | Yes | Yes |
| Geographic controls | No | No | Yes | No | No | Yes |
| $R^2$ | 0.598 | 0.829 | 0.849 | 0.598 | 0.829 | 0.849 |
| Observations | 660 | 660 | 660 | 660 | 660 | 660 |

Notes: Estimation method is OLS using data from Dincecco et al. (2022). Unit of analysis is district. The dependent variable is ln(0.01 + Luminosity) averaged between 1992 and 2010. Variable of interest is pre-colonial exposure to land battles between 1000 and 1757. Geographic controls include latitude, longitude, altitude, ruggedness, precipitation, land quality, dry rice suitability, wet rice suitability, wheat suitability and malaria risk. Population density is ln(Population Density) in 1990. Robust SEs in parenthesis are calculated using the robust command in Stata, and using White HC1 standard errors in R. Significant at the ***[1%] **[5%] *[10%] level.

Table 2: *Pre-Colonial Conflict and Economic Development: IV*

| | Original Study | | | R reproduction | | |
|---|---|---|---|---|---|---|
| *Panel A: First stage* | | | | | | |
| *Dependent variable:* | Pre-colonial conflict exposure | | | | | |
| | (1) | (2) | (3) | (1′) | (2′) | (3′) |
| Proximity to Khyber Pass | 0.204*** | 0.094*** | 0.080*** | 0.204*** | 0.094*** | 0.080*** |
| | (0.018) | (0.025) | (0.024) | (0.018) | (0.025) | (0.024) |
| Population density | Yes | Yes | Yes | Yes | Yes | Yes |
| State FE | No | Yes | Yes | No | Yes | Yes |
| Geographic controls | No | No | Yes | No | No | Yes |
| $R^2$ | 0.415 | 0.645 | 0.665 | 0.415 | 0.593 | 0.645 |
| Observations | 660 | 660 | 660 | 660 | 660 | 660 |
| *Panel B: Second stage* | | | | | | |
| *Dependent variable:* | ln(0.01 + Luminosity) | | | | | |
| | (1) | (2) | (3) | (1′) | (2′) | (3′) |
| Pre-colonial conflict exposure | 4.930*** | 4.626*** | 3.482** | 4.930*** | 4.626*** | 3.482** |
| | (0.609) | (1.291) | (1.389) | (0.607) | (1.328) | (1.441) |
| Population density | Yes | Yes | Yes | Yes | Yes | Yes |
| State FE | No | Yes | Yes | No | Yes | Yes |
| Geographic controls | No | No | Yes | No | No | Yes |
| $\mathrm{R}^2$ | 0.593 | 0.814 | 0.843 | 0.593 | 0.814 | 0.843 |
| Observations | 660 | 660 | 660 | 660 | 660 | 660 |

Notes: Estimation method is 2SLS using data from Dincecco et al. (2022). Unit of analysis is district. In Panel A (first stage) the dependent variable is pre-colonial conflict exposure to land battles between 1000 and 1757, the variable of interest is proximity to the Khyber pass. In panel B (second stage) the dependent variable is ln(0.01 + Luminosity) averaged between 1992 and 2010. Variable of interest is pre-colonial exposure to land battles between 1000 and 1757. Geographic controls include latitude, longitude, altitude, ruggedness, precipitation, land quality, dry rice suitability, wet rice suitability, wheat suitability and malaria risk. Population density is ln(Population Density) in 1990. Robust SEs in parenthesis are calculated using the robust command in Stata, and using White HC1 standard errors in R. Significant at the ***[1%] **[5%] *[10%] level.

Table 3: *Pre-Colonial Conflict and Economic Development: Alternative Conflict Proxies*

| Dependent Variable | $\ln(0.01 + \text{Luminosity})$ | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Land battles (original study) | 3.713*** | 1.601*** | 1.465*** |
| | (0.305) | (0.380) | (0.370) |
| All conflicts | 3.009*** | 0.761*** | 0.681*** |
| | (0.278) | (0.256) | (0.250) |
| Multi-day | 4.761*** | 0.612 | 0.551 |
| | (1.678) | (0.565) | (0.496) |
| Multi-year | 33.283*** | 2.565 | -0.324 |
| | (5.131) | (3.466) | (3.462) |
| Naval | 34.771*** | -3.893 | -5.511 |
| | (9.619) | (6.608) | (6.326) |
| One-day | 4.124*** | 1.324*** | 1.256*** |
| | (0.384) | (0.350) | (0.358) |
| Sacking, Razing | 16.887*** | 4.244*** | 4.796*** |
| | (3.158) | (1.638) | (1.697) |
| Siege | 5.104*** | 0.339 | 0.320 |
| | (1.660) | (0.462) | (0.422) |
| Storming | 4.670* | -0.868 | -0.519 |
| | (2.781) | (0.729) | (0.708) |
| Population density | Yes | Yes | Yes |
| State FE | No | Yes | Yes |
| Geographic controls | No | No | Yes |
| Observations | 660 | 660 | 660 |

Notes: Estimation method is OLS using data from Dincecco et al. (2022). Alternative proxies for conflict are provided by the authors in the original dataset. Unit of analysis is district. The dependent variable is $\ln(0.01 + \text{Luminosity})$ averaged between 1992 and 2010. Variable of interest is pre-colonial exposure to land battles between 1000 and 1757. Geographic controls include latitude, longitude, altitude, ruggedness, precipitation, land quality, dry rice suitability, wet rice suitability, wheat suitability and malaria risk. Population density is ln(Population Density) in 1990. Robust SEs in parenthesis are calculated using the robust command in Stata. Significant at the ***[1%] **[5%] *[10%] level.

Table 4: *Pre-Colonial Conflict and Economic Development: Alternative Conflict Proxies*

| *IV Second stage* | | | |
|---|---|---|---|
| *Dependent variable:* | $\ln(0.01 + \text{Luminosity})$ | | |
| | (1) | (2) | (3) |
| Land battles (original study) | 4.930*** | 4.626*** | 3.482** |
| | (0.609) | (1.291) | (1.389) |
| All conflicts | 4.091*** | 3.443*** | 2.791** |
| | (0.511) | (0.940) | (1.094) |
| Multi-day | 24.673*** | 117.641 | -69.692 |
| | (5.366) | (338.930) | (226.214) |
| Multi-year | 683.362* | -284.866 | -98.369 |
| | (401.619) | (339.436) | (86.406) |
| Naval | -457.864*** | -3344.237* | -1599.681 |
| | (91.611) | (1960.296) | (3093.860) |
| One-day | 4.940*** | 3.503*** | 2.612*** |
| | (0.665) | (0.817) | (0.978) |
| Sacking, Razing | 20.476*** | 13.684*** | 8.862*** |
| | (2.872) | (2.853) | (3.344) |
| Siege | 22.834*** | 13.407*** | 13.943* |
| | (4.777) | (4.456) | (7.691) |
| Storming | -342.058 | 84.076 | -156.537 |
| | (446.155) | (101.599) | (582.155) |
| Population density | Yes | Yes | Yes |
| State FE | No | Yes | Yes |
| Geographic controls | No | No | Yes |
| Observations | 660 | 660 | 660 |

Notes: Estimation method is 2SLS, showing second stage only using data from Dincecco et al. (2022). Alternative proxies for conflict are provided by the authors in the original dataset. Unit of analysis is district. The dependent variable is $\ln(0.01 + \text{Luminosity})$ averaged between 1992 and 2010. Variable of interest is pre-colonial exposure to conflict between 1000 and 1757. Geographic controls include latitude, longitude, altitude, ruggedness, precipitation, land quality, dry rice suitability, wet rice suitability, wheat suitability and malaria risk. Population density is ln(Population Density) in 1990. Robust SEs in parenthesis are calculated using the robust command in Stata. Significant at the ***[1%] **[5%] *[10%] level.

Table 5: *Pre-Colonial Conflict and Economic Development: Alternative Measures of Conflict Exposure*

| Dependent Variable | ln(0.01 + Luminosity) | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| A: Land battles 1001-1757 | 3.713*** | 1.601*** | 1.465*** |
| within 250km (original study) | (0.305) | (0.380) | (0.370) |
| $R^2$ | 0.598 | 0.829 | 0.849 |
| B: Land battles 1001-1757 | 3.553*** | 0.728*** | 0.626** |
| within 250km (replication using HCED) | (0.310) | (0.278) | (0.276) |
| $R^2$ | 0.627 | 0.826 | 0.846 |
| B': Conflicts 1001-1757 | 3.004*** | 0.756*** | 0.678*** |
| within 250km (replication using DFMM) | (0.278) | (0.255) | (0.249) |
| $R^2$ | 0.619 | 0.827 | 0.847 |
| C: Land battles 1001-1757 | 2.586*** | 0.559** | 0.515** |
| within 250km, 1 omitted | (0.359) | (0.254) | (0.239) |
| $R^2$ | 0.610 | 0.826 | 0.847 |
| D: Land battles 1001-1757 | 1.967*** | 0.525*** | 0.467*** |
| within 155.343 miles | (0.160) | (0.157) | (0.157) |
| $R^2$ | 0.622 | 0.827 | 0.847 |
| E: Land battles 1001-1757 | 0.089*** | 0.036*** | 0.031*** |
| within 250km, in 100km | (0.005) | (0.006) | (0.007) |
| $R^2$ | 0.644 | 0.830 | 0.849 |
| F: Land battles 1001-1757 | 0.127*** | 0.046*** | 0.040*** |
| within 250km, Gaussian | (0.007) | (0.009) | (0.009) |
| $R^2$ | 0.635 | 0.829 | 0.848 |
| G: Land battles 1001-1757 | 0.040*** | 0.016*** | 0.014*** |
| within 250km, count | (0.002) | (0.003) | (0.003) |
| $R^2$ | 0.647 | 0.830 | 0.849 |
| Population density | Yes | Yes | Yes |
| State FE | No | Yes | Yes |
| Geographic controls | No | No | Yes |
| Observations | 660 | 660 | 660 |

Notes: Estimation method is OLS. Variable of interest is pre-colonial exposure to land battles between 1000 and 1757 within 250km of the district centroid, with alternative measurement as follows: A: Estimates reported by DFMM. B: Replication using land battles data from Historical Conflict Event Dataset (Miller and Bakar 2023), code written by the current authors. B′: Replication using conflict data from Dincecco et al. (2022), code written by the current authors. C: Distance measure calculated without adding one to $distance_{i,c}$ before taking the inverse. D: Distance measured in miles, land battles within 155.343 miles of district centroid. E: Distance re-scaled measured in 100km before taking inverse. F: Gaussian distance measure, $\sum_{c \in C} e^{-distance_{i,c}}$ G: Count of land battles within 250km. Distance Unit of analysis is district. The dependent variable is ln(0.01 + Luminosity) averaged between 1992 and 2010. Geographic controls include latitude, longitude, altitude, ruggedness, precipitation, land quality, dry rice suitability, wet rice suitability, wheat suitability and malaria risk. Population density is ln(Population Density) in 1990. Robust SEs in parenthesis are calculated using the robust command in Stata. Significant at the ***[1%] **[5%] *[10%] level.

Table 6: *Pre-Colonial Conflict and Economic Development: IV Second Stage with alternative measures of distance to conflict exposure*

| *IV Second stage* | | | |
|---|---|---|---|
| *Dependent variable:* | $\ln(0.01 + \text{Luminosity})$ | | |
| | (1) | (2) | (3) |
| A: Pre-colonial conflict exposure | 4.930*** | 4.626*** | 3.482** |
| (original study) | (0.609) | (1.291) | (1.389) |
| $\mathrm{R}^2$ | 0.593 | 0.814 | 0.843 |
| B: Pre-colonial conflict exposure | 4.123*** | 3.834*** | 3.183** |
| (replication using HCED) | (0.504) | (0.970) | (1.272) |
| $R^2$ | 0.609 | 0.799 | 0.831 |
| B′: Pre-colonial conflict exposure | 4.098*** | 3.434*** | 2.773** |
| (replication using data from DFMM) | (0.515) | (0.938) | (1.085) |
| $R^2$ | 0.609 | 0.799 | 0.831 |
| C: Pre-colonial conflict exposure | 3.955*** | 3.697*** | 3.071** |
| (1 omitted) | (0.515) | (0.945) | (1.256) |
| $R^2$ | 0.617 | 0.795 | 0.827 |
| D: Pre-colonial conflict exposure | 2.616*** | 2.425*** | 2.013** |
| (distance units: miles) | (0.311) | (0.610) | (0.797) |
| $R^2$ | 0.629 | 0.805 | 0.833 |
| E: Pre-colonial conflict exposure | 0.099*** | 0.085*** | 0.069*** |
| (distance units: 100km) | (0.009) | (0.018) | (0.024) |
| $R^2$ | 0.652 | 0.824 | 0.846 |
| F: Pre-colonial conflict exposure | 0.149*** | 0.130*** | 0.105*** |
| (Gaussian) | (0.015) | (0.030) | (0.037) |
| $R^2$ | 0.644 | 0.820 | 0.843 |
| G: Pre-colonial conflict exposure | 0.044*** | 0.037*** | 0.031*** |
| (Count) | (0.004) | (0.007) | (0.011) |
| $R^2$ | 0.657 | 0.823 | 0.845 |
| Population density | Yes | Yes | Yes |
| State FE | No | Yes | Yes |
| Geographic controls | No | No | Yes |
| Observations | 660 | 660 | 660 |

Notes: Estimation method is 2SLS, showing second stage only. Unit of analysis is district. Second stage dependent variable is $\ln(0.01 + \text{Luminosity})$ averaged between 1992 and 2010. Variable of interest is pre-colonial exposure to land battles between 1000 and 1757 with alternative measures as follows. A: Estimates reported by DFMM. B: Replication using crosswalk and data provided by DFMM, code written by the current authors. C: Distance measure calculated without adding one to $distance_{i,c}$ before taking the inverse. D: Distance measured in miles, land battles within 155.343 miles of district centroid. E: Distance re-scaled measured in 100km before taking inverse. F: Gaussian distance measure, $\sum_{c \in C} e^{-distance_{i,c}}$ G: Count of land battles within 250km. Geographic controls include latitude, longitude, altitude, ruggedness, precipitation, land quality, dry rice suitability, wet rice suitability, wheat suitability and malaria risk. Population density is ln(Population Density) in 1990. Robust SEs in parenthesis are calculated using the robust command in Stata. Significant at the ***[1%] **[5%] *[10%] level.

Table 7: *Count of Conflicts by Time Period*

| Century | N, using DFMM | N, using HCED |
|---|---|---|
| 1000 - 1100 | 12 | 11 |
| 1101 - 1200 | 5 | 5 |
| 1201 - 1300 | 6 | 9 |
| 1301 - 1400 | 12 | 13 |
| 1401 - 1500 | 4 | 4 |
| 1501 - 1600 | 47 | 39 |
| 1601 - 1700 | 64 | 65 |
| 1701 - 1757 | 96 | 91 |
| Total 1000-1757 | 246 | 237 |

Notes: Number of conflicts by century using data from Dincecco et al. (2022) (N, using DFMM) and the Historical Conflict Event Dataset (HCED, Miller and Bakar (2023)). Calculations by the present authors.

Table 8: *Descriptive Statistics of Conflict Exposure by Time Period*

| Century | N | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| 1000 - 1100 | 666 | 0.003 | 0.007 | 0 | 0.11 |
| 1101 - 1200 | 666 | 0.004 | 0.001 | 0 | 0.15 |
| 1201 - 1300 | 666 | 0.004 | 0.014 | 0 | 0.25 |
| 1301 - 1400 | 666 | 0.003 | 0.011 | 0 | 0.20 |
| 1401 - 1500 | 666 | 0 | 0 | 0 | 0 |
| 1501 - 1600 | 666 | 0.012 | 0.02 | 0 | 0.25 |
| 1601 - 1700 | 666 | 0.016 | 0.027 | 0 | 0.44 |
| 1701 - 1757 | 666 | 0.06 | 0.07 | 0 | 0.41 |
| Total 1000-1757 | 666 | 0.07 | 0.01 | 0 | 0.61 |

Notes: Mean of conflict exposure to Land Battles from 1001-1757 within 250lm of district centroid calculated over districts by century using data from Dincecco et al. (2022). Calculation by the present authors.

Table 9: *Pre-Colonial Conflict and Economic Development: Main Results using different Time Periods*

| *Dependent Variable*: | ln(0.01 + Luminosity) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Pre-colonial conflict exposure | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| 1000-1757 (original study) | 1.465*** | | | | | | | |
| | (0.370) | | | | | | | |
| 1000-1100 (replication) | | 11.05*** | | | | | | |
| | | (3.732) | | | | | | |
| 1101-1200 (replication) | | | 7.779*** | | | | | |
| | | | (2.665) | | | | | |
| 1201-1300 (replication) | | | | 5.502* | | | | |
| | | | | (3.100) | | | | |
| 1301-1400 (replication) | | | | | 4.206* | | | |
| | | | | | (2.430) | | | |
| 1401-1500 (replication) | | | | | | - | | |
| 1501-1600 (replication) | | | | | | | 2.027* | |
| | | | | | | | (1.039) | |
| 1601-1700 (replication) | | | | | | | | 3.040*** |
| | | | | | | | | (0.860) |
| Population density | Yes | Yes | Yes | Yes | Yes | - | Yes | Yes |
| State FE | Yes | Yes | Yes | Yes | Yes | - | Yes | Yes |
| Geographic controls | Yes | Yes | Yes | Yes | Yes | - | Yes | Yes |
| Observations | 660 | 660 | 660 | 660 | 660 | - | 660 | 660 |
| R-squared | 0.847 | 0.847 | 0.847 | 0.847 | 0.846 | - | 0.846 | 0.847 |

Notes: Estimation method is OLS using data from Dincecco et al. (2022). Unit of analysis is district. The dependent variable is ln(0.01 + Luminosity) averaged between 1992 and 2010. Variable of interest is pre-colonial exposure to land battles between 1000 and 1757, with rows indicating the time period used in each specification in 100 year blocks. Conflict exposure during the period 1400-1500 is omitted as no land battles occured during this time in the data. Geographic controls include latitude, longitude, altitude, ruggedness, precipitation, land quality, dry rice suitability, wet rice suitability, wheat suitability and malaria risk. Population density is ln(Population Density) in 1990. Robust SEs in parenthesis are calculated using the robust command in Stata. Significant at the ***[1%] **[5%] *[10%] level.

Table 10: *Pre-Colonial Conflict and Economic Development: IV with Different Time Periods*

| *Dependent Variable:* | ln(0.01 + Luminosity) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Pre-colonial conflict exposure | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| 1000-1757 (original study) | 3.482** | | | | | | | |
| | (1.441) | | | | | | | |
| 1000-1100 (replication) | | 21.06** | | | | | | |
| | | (8.216) | | | | | | |
| 1101-1200 (replication) | | | 25.73** | | | | | |
| | | | (12.42) | | | | | |
| 1201-1300 (replication) | | | | 14.25** | | | | |
| | | | | (6.207) | | | | |
| 1301-1400 (replication) | | | | | 60.53* | | | |
| | | | | | (32.91) | | | |
| 1401-1500 (replication) | | | | | | - | | |
| 1501-1600 (replication) | | | | | | | 12.34** | |
| | | | | | | | (5.395) | |
| 1601-1700 (replication) | | | | | | | | -45.57 |
| | | | | | | | | (54.26) |
| Population density | Yes | Yes | Yes | Yes | Yes | - | Yes | Yes |
| State FE | Yes | Yes | Yes | Yes | Yes | - | Yes | Yes |
| Geographic controls | Yes | Yes | Yes | Yes | Yes | - | Yes | Yes |
| Observations | 660 | 660 | 660 | 660 | 660 | - | 660 | 660 |
| R-squared | 0.849 | 0.846 | 0.838 | 0.843 | 0.731 | - | 0.835 | 0.376 |

Notes: Estimation method is 2SLS using data from Dincecco et al. (2022), showing second stage only. Unit of analysis is district. The dependent variable is ln(0.01+Luminosity) averaged between 1992 and 2010. Variable of interest is pre-colonial exposure to land battles between 1000 and 1757, reported broken down into 100 year periods. Geographic controls include latitude, longitude, altitude, ruggedness, precipitation, land quality, dry rice suitability, wet rice suitability, wheat suitability and malaria risk. Population density is ln(Population Density) in 1990. Robust SEs in parenthesis are calculated using the robust command in Stata. Significant at the ***[1%] **[5%] *[10%] level.

Table 11: *Pre-Colonial Conflict and Pre-colonial-Era State-Making*

| *Dependent variable:* | Important Mughal sites | State history | | |
|---|---|---|---|---|
| | | Babur | Akbar | Aurangzeb |
| Pre-colonial conflict exposure | (1) | (2) | (3) | (4) |
| 1000-1757 (original study) | 0.954* | | | |
| | (0.497) | | | |
| 1000-1526 (original study) | | 0.513** | | |
| | | (0.229) | | |
| 1000-1556 (original study) | | | 0.723*** | |
| | | | (0.262) | |
| 1000-1658 (original study) | | | | -0.080 |
| | | | | (0.173) |
| $R^2$ | 0.122 | 0.768 | 0.715 | 0.718 |
| Observations | 659 | 659 | 659 | 659 |
| 1500-1757 (replication) | 1.241** | 0.552*** | 0.292 | 0.719*** |
| | (0.590) | (0.162) | (0.206) | (0.170) |
| $R^2$ | 0.124 | 0.771 | 0.714 | 0.726 |
| Observations | 659 | 659 | 659 | 659 |
| Population density | Yes | Yes | Yes | Yes |
| State FE | Yes | Yes | Yes | Yes |
| Geographic controls | Yes | Yes | Yes | Yes |

Notes: Estimation method is OLS using data from Dincecco et al. (2022). Unit of analysis is district. Dependent variable in column 1 is number of important Mughal-era sites including public works. Dependent variables in columns 2–4 are state longevity in terms of districts incorporated into the Mughal Empire by Babur (1526–30), Akbar (1556–1605) and Aurangzeb (1658–1707). Variable of interest is pre-colonial conflict exposure to land battles. It spans 1000-1757, 1000-1526, 1000-1556 and 1000-1658 in the original study and 1500–1757 in the replication. Geographic controls include latitude, longitude, altitude, ruggedness, precipitation, land quality, dry rice suitability, wet rice suitability, wheat suitability and malaria risk. Population density is ln(PopulationDensity) in 1500. Robust SEs in parenthesis are calculated using the robust command in Stata. Significant at the ***[1%] **[5%] *[10%] level.

Table 12: *Pre-Colonial Conflict and Colonial Fiscal Development*

| *Dependent variable:* | 1881 | | | | | | 1931 | |
|---|---|---|---|---|---|---|---|---|
| | All | | British India | | Princely States | | British India | |
| Pre-colonial conflict exposure | (1) ln(Tax/Area) | (2) ln(Tax/Person) | (3) ln(Tax/Area) | (4) ln(Tax/Person) | (5) ln(Tax/Area) | (6) ln(Tax/Person) | (7) ln(Tax/Acre) | (8) ln(Tax/Person) |
| 1000-1757 (original study) | 2.246*** | 1.245*** | 2.208*** | 1.157*** | 6.386* | 1.612 | 1.310 | 0.835 |
| | (0.550) | (0.382) | (0.516) | (0.354) | (3.524) | (2.228) | (0.911) | (0.705) |
| 1500-1757 (replication) | 2.335*** | 1.304*** | 2.173*** | 1.119*** | 10.41*** | 3.391 | 1.298 | 0.757 |
| | (0.659) | (0.446) | (0.589) | (0.396) | (3.766) | (2.585) | (1.120) | (0.794) |
| Population density | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| State FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Geographic controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Standardized beta coefficient | 0.256 | 0.173 | 0.281 | 0.227 | 0.496 | 0.210 | 0.135 | 0.098 |
| $R^2$ | 0.468 | 0.518 | 0.596 | 0.545 | 0.606 | 0.408 | 0.731 | 0.696 |
| Observations | 270 | 274 | 200 | 200 | 70 | 74 | 145 | 144 |

Notes: Estimation method is OLS using data from Dincecco et al. (2022). Unit of analysis is district. Dependent variables are as follows. ln($Tax/Area$), 1881 and ln($Tax/Person$), 1881 measures land revenue in 1,000 rupees per square kilometre or per capita, in 1881 for districts under direct British rule and/or indirect rule (i.e., major Princely states). ln($Tax/Acre$), 1931 and ln($Tax/Person$), 1931 measures average land revenue in rupees per acre or per capita, in 1931 for districts in British India. Variable of interest is pre-colonial conflict exposure to land battles between 1000 and 1757. Geographic controls include latitude, longitude, altitude, ruggedness, precipitation, land quality, dry rice suitability, wet rice suitability, wheat suitability and malaria risk. Population density is ln($PopulationDensity$) in 1850 in columns 1–6 and 1930 in columns 7–8. Robust SEs in parenthesis are calculated using the robust command in Stata. Significant at the ***[1%] **[5%] *[10%] level.

Table 13: *Pre-Colonial Conflict Versus Colonial and Post-Colonial Conflict*

| *Dependent variable:* | Colonial conflict exposure 1758-1839 | | Colonial conflict exposure 1840-1946 | | Post-colonial conflict exposure 1947-2010 | |
|---|---|---|---|---|---|---|
| Pre-colonial conflict exposure | Land battles (1) | All conflicts (2) | Land battles (3) | All conflicts (4) | Land battles (5) | All conflicts (6) |
| 1000-1757 (original study) | 0.170*** | 0.446*** | 0.037 | 0.313 | -0.024*** | -0.030*** |
| | (0.036) | (0.090) | (0.039) | (0.306) | (0.005) | (0.007) |
| 1500-1757 (replication) | 0.201*** | 0.565*** | 0.101** | 0.618 | -0.0250*** | -0.0288*** |
| | (0.0497) | (0.122) | (0.0497) | (0.469) | (0.00632) | (0.00847) |
| Population density | Yes | Yes | Yes | Yes | Yes | Yes |
| State FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Geographic controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 660 | 660 | 660 | 660 | 660 | 660 |

Notes: Estimation method is OLS using data from Dincecco et al. (2022). Unit of analysis is district. Dependent variable is colonial conflict exposure to land battles between 1758 and 1839 in column 1 and to all conflict types in column 2. Similarly, it is colonial conflict exposure between 1840 and 1946 in columns 3–4, and post-colonial conflict exposure between 1947 and 2010 in columns 5– 6. Variable of interest is pre-colonial conflict exposure to land battles between 1500 and 1757. Geographic controls include latitude, longitude, altitude, ruggedness, precipitation, land quality, dry rice suitability, wet rice suitability, wheat suitability and malaria risk. Population density is ln(PopulationDensity) in 1750 in columns 1–2, in 1850 in columns 3–4 and in 1950 in columns 5–6. Robust SEs in parenthesis are calculated using the robust command in Stata. Significant at the ***[1%] **[5%] *[10%] level.

Table 14: *Pre-Colonial Conflict and Post-Colonial Political Violence*

| *Dependent variable:* | Political Violence | Maoist Control | Fractionalisation | |
|---|---|---|---|---|
| Pre-colonial conflict exposure | (1) | (2) | Linguistic (3) | Religious (4) |
| 1000-1757 (original study) | -0.241** | -0.381** | -0.209* | 0.080 |
| | (0.102) | (0.163) | (0.113) | (0.071) |
| $R^2$ | 0.408 | 0.281 | 0.570 | 0.557 |
| Observations | 660 | 395 | 660 | 660 |
| 1500-1757 (replication) | -0.329** | -0.276* | -0.210 | 0.0258 |
| | (0.133) | (0.162) | (0.148) | (0.0891) |
| $R^2$ | 0.409 | 0.191 | 0.570 | 0.556 |
| Observations | 660 | 293 | 660 | 660 |
| Population density | Yes | Yes | Yes | Yes |
| State FE | Yes | Yes | Yes | Yes |
| Geographic controls | Yes | Yes | Yes | Yes |

Notes: Estimation method is OLS using data from Dincecco et al. (2022). Unit of analysis is district. Dependent variable in column 1 is Political violence, defined as fatalities per district between 2015 and 2018 (in hundreds). Dependent variable in column 2 is Linguistic Fractionalisation, defined as 1 minus the Herfindahl index of language population shares in 2001. Dependent variable in column 3 is Religious Fractionalization, defined as 1 minus the Herfindahl index of religion population shares in 2001. Variable of interest is pre-colonial conflict exposure to land battles between 1500 and 1757. Geographic controls include latitude, longitude, altitude, ruggedness, precipitation, land quality, dry rice suitability, wet rice suitability, wheat suitability and malaria risk. Population density is ln(PopulationDensity) in 1990. Robust SEs in parenthesis are calculated using the robust command in Stata. Significant at the ***[1%] **[5%] *[10%] level.

Table 15: *Pre-Colonial Conflict and Irrigation Infrastructure*

| *Dependent Variable:* | % Irrigated | | ln(Yield) | % Non-Agriculture |
|---|---|---|---|---|
| | 1931 | 1956-87 | | |
| Pre-colonial conflict exposure | (1) | (2) | (3) | (4) |
| 1000-1757 (original study) | 21.275** | 37.413** | 0.737* | 0.197** |
| | (10.357) | (15.758) | (0.381) | (0.085) |
| $R^2$ | 0.391 | 0.611 | 0.683 | 0.566 |
| Observations | 257 | 271 | 271 | 660 |
| 1500-1757 (replication) | 15.81 | 39.40 | 0.168 | 0.249** |
| | (10.47) | (24.95) | (0.621) | (0.107) |
| Constant | 17.27 | -23.88 | 0.860 | 0.416 |
| | (29.93) | (65.80) | (1.725) | (0.341) |
| $R^2$ | 0.382 | 0.641 | 0.728 | 0.567 |
| Observations | 257 | 208 | 208 | 660 |
| Population density | Yes | Yes | Yes | Yes |
| State FE | Yes | Yes | Yes | Yes |
| Geographic controls | Yes | Yes | Yes | Yes |

Notes: Estimation method is OLS using data from Dincecco et al. (2022). Unit of analysis is district. Dependent variables are as follows: %Irrigated measures the proportion of area sown with canal irrigation in 1931 (column 1) and the proportion of gross cropped area that is irrigated averaged between 1956 and 1987 (column 2); ln(Yield) measures the total yield across 15 major crops averaged between 1956 and 1987 (column 3); and %Non-agriculture measures the share of non-agricultural workers in 2011 (column 4). Variable of interest is pre-colonial conflict exposure to land battles between 1000 and 1757. Geographic controls include latitude, longitude, altitude, ruggedness, precipitation, land quality, dry rice suitability, wet rice suitability, wheat suitability and malaria risk. Population density is ln(PopulationDensity) in 1900 in column 1, in 1950 in columns 2–3, and 1990 in column 4. Note that the number of observations in columns 2 and 3 differ between the original study and the replication due to data unavailability, see section 3.3 for a discussion. Robust SEs in parenthesis are calculated using the robust command in Stata. Significant at the ***[1%] **[5%] *[10%] level.

Table 16: *Pre-Colonial Conflict and Literacy*

| *Dependent variable:* | %Literacy | | | |
|---|---|---|---|---|
| | 1881 | 1921 | 1961-91 | 2011 |
| | (1) | (2) | (3) | (4) |
| 1000-1757 (original study) | -1.933 | -5.635 | 11.796* | 10.146** |
| | (3.188) | (3.772) | (6.888) | (4.119) |
| $R^2$ | 0.464 | 0.556 | 0.623 | 0.599 |
| Observations | 251 | 303 | 271 | 626 |
| 1500-1757 (replication) | -0.199 | -4.827 | 12.19 | 1.615 |
| | (3.529) | (5.129) | (8.756) | (6.119) |
| $R^2$ | 0.463 | 0.554 | 0.614 | 0.599 |
| Observations | 251 | 303 | 264 | 541 |
| Population density | Yes | Yes | Yes | Yes |
| State FE | Yes | Yes | Yes | Yes |
| Geographic controls | Yes | Yes | Yes | Yes |

Notes: Estimation method is OLS using data from Dincecco et al. (2022). Unit of analysis is district. Dependent variables are as follows: %Literacy, 1881 is the proportion of 'literate' persons in 1881; %Literacy, 1921 is the proportion of persons that can read and write in 1921; %Literacy, 1961–91 is the literacy rate averaged between 1961 and 1991; and %Literacy, 2011 measures the adult literacy rate across both rural and urban populations for ages 7-plus. Variable of interest is pre-colonial conflict exposure to land battles between 1000 and 1757 (top panel, direct replication of original study) and exposure to land battles between 1500 and 1757. Geographic controls include latitude, longitude, altitude, ruggedness, precipitation, land quality, dry rice suitability, wet rice suitability, wheat suitability and malaria risk. Population density is ln(PopulationDensity) in 1850 in column 1, in 1900 in column 2, in 1950 in column 3 and in 2011 in column 4. Observations differ between the direct replication (top) panel and the lower panel due to missing data in the replication package, see section 3.3 for a discussion. Robust SEs in parenthesis are calculated using the robust command in Stata. Significant at the ***[1%] **[5%] *[10%] level.

Table 17: *Pre-Colonial Conflict and Education*

| *Dependent variable:* | %Primary | %High | %InfantMortality |
|---|---|---|---|
| Pre-colonial conflict exposure | (1) | (2) | (3) |
| 1000-1757 (original study) | 18.683* | -16.094** | -35.283** |
| | (11.150) | (6.553) | (14.405) |
| $R^2$ | 0.712 | 0.840 | 0.674 |
| Observations | 203 | 187 | 270 |
| 1500-1757 (replication) | 32.982** | -6.491 | -33.871 |
| | (13.316) | (4.163) | (21.253) |
| $R^2$ | 0.823 | 0.891 | 0.688 |
| Observations | 157 | 147 | 208 |
| Population density | Yes | Yes | Yes |
| State FE | Yes | Yes | Yes |
| Geographic controls | Yes | Yes | Yes |

Notes: Estimation method is OLS using data from Dincecco et al. (2022). Unit of analysis is district. Dependent variables are as follows: %Primary measures the proportion of villages having a primary school in 1981; %High measure the proportion of villages having a high school in 1981; and %InfantMortality is the infant mortality rate in 1991. Variable of interest is pre-colonial conflict exposure to land battles between 1000 and 1757. Geographic controls include latitude, longitude, altitude, ruggedness, precipitation, land quality, dry rice suitability, wet rice suitability, wheat suitability, and malaria risk. Population density is ln(PopulationDensity) in 1950 in columns 1 and 2, and in 1990 in column 3. Observations differ between the direct replication (top) panel and the lower panel due to missing data in the replication package, see section 3.3 for a discussion. Robust SEs in parenthesis are calculated using the robust command in Stata. Significant at the ***[1%] **[5%] *[10%] level.

Table 18: *Pre-Colonial Conflict and Post Colonial Political Violence*

| | Original Study | Replication using different data | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Time period | 2015-2018 | 2001-2021 | 2015-2018 |
| Pre-colonial conflict exposure | -0.241** | -0.639* | -0.034 |
| | (0.102) | (0.349) | (0.064) |
| Population density | Yes | Yes | Yes |
| State FE | Yes | Yes | Yes |
| Geographic controls | Yes | Yes | Yes |
| $R^2$ | 0.408 | 0.447 | 0.258 |
| Observations | 660 | 660 | 660 |

Notes: Estimation method is OLS using data from Dincecco et al. (2022). Unit of analysis is district.
The dependent variable in column (1) is political violence, defined as fatalities per district between 2015 and 2018 (in hundreds) using ACLED data. Column (1) replicates the results of Dincecco et al. (2022).
The dependent variable in column (2) and (3) is organised violence, defined as fatalities per state (in hundreds) based on the UCDP data. Column (2) focus on the time from 2001 to 2021 and column (3) on the time from 2015 until 2018. The variable of interest is pre-colonial conflict exposure to land battles between 1000 and 1757. Geographic controls include latitude, longitude, altitude, ruggedness, precipitation, land quality, dry rice suitability, wet rice suitability, wheat suitability and malaria risk. Population density is ln(PopulationDensity) in 1990. Robust SEs in parenthesis are calculated using the robust command in Stata. Significant at the ***[1%] **[5%] *[10%] level.

Table 19: *Reproducibility and Replicability Results*

| *Replication Type*: | Sub-type | Test | N tests | N reproduced/replicated | % |
|---|---|---|---|---|---|
| Computational | Reproduction | Table 1 | 3 | 3 | 100 |
| | | Table 2 | 3 | 3 | 100 |
| | | First Digits (Figure 1) | 19 | 19 | 100 |
| **Total Computational** | | | **25** | **25** | **100** |
| Direct | Alternative conflict proxies, tables 3 - 4 | All conflicts | 6 | 6 | 100 |
| | | Multi-day | 6 | 2 | $33.\overline{33}$ |
| | | Multi-year | 6 | 2 | $33.\overline{33}$ |
| | | Naval | 6 | 1 | $16.\overline{66}$ |
| | | One-day | 6 | 6 | 100 |
| | | Sacking | 6 | 6 | 100 |
| | | Seige | 6 | 5 | $83.\overline{33}$ |
| | | Storming | 6 | 1 | $16.\overline{66}$ |
| | *Total, alternative conflict proxies* | | 48 | 29 | 60.42 |
| | Alternative measures of conflict exposure, tables 5 - 6 | Replication | 2 | 2 | 100 |
| | | 1 omitted | 2 | 2 | 100 |
| | | Distance unit: miles | 2 | 2 | 100 |
| | | Distance unit: 100km | 2 | 2 | 100 |
| | | Gaussian | 2 | 2 | 100 |
| | | Count | 2 | 2 | 100 |
| | *Total, alternative conflict exposure measures* | | 12 | 12 | 100 |
| | Different time periods, table 9 | 1000-1100 | 1 | 1 | 100 |
| | | 1101-1200 | 1 | 1 | 100 |
| | | 1201-1300 | 1 | 1 | 100 |
| | | 1301-1400 | 1 | 1 | 100 |
| | | 1501-1600 | 1 | 1 | 100 |
| | | 1601-1700 | 1 | 1 | 100 |
| | | Total | 6 | 6 | 100 |
| | Different time periods, table 10 | 1000-1100 | 1 | 1 | 100 |
| | | 1101-1200 | 1 | 1 | 100 |
| | | 1201-1300 | 1 | 1 | 100 |
| | | 1301-1400 | 1 | 1 | 100 |
| | | 1501-1600 | 1 | 1 | 100 |
| | | 1601-1700 | 1 | 0 | 0 |
| | | Total | 6 | 5 | $83.\overline{33}$ |
| | *Total, different time periods (main results)* | | 12 | 11 | 91.67 |
| | Different time periods | Table 11 | 4 | 2 | 50 |
| | | Table 12 | 8 | 8 | 100 |
| | | Table 13 | 6 | 5 | $83.\overline{33}$ |
| | | Table 14 | 4 | 3 | 75 |
| | | Table 15 | 4 | 1 | 25 |
| | | Table 16 | 4 | 2 | 50 |
| | | Table 17 | 3 | 1 | $33.\overline{33}$ |
| | | Total | 33 | 22 | $66.\overline{66}$ |
| | *Total, different time periods (mechanisms results)* | | 33 | 22 | $66.\overline{67}$ |
| | Alternative Political Violence Data | Table 18 | 2 | 1 | 50 |
| | | Total | 2 | 1 | 50 |
| **Total Direct** | | | **107** | **75** | **70.1** |
| **Grand Total** | | | **132** | **100** | **75.76** |

Notes: We define a positive replication as an estimate of the same sign (positive/negative) and significance (significant/not significantly different from zero) as that reported the original paper. This definition of course precludes difference in the magnitudes of estimates, which we discuss in the main text.

# References


Azevedo CDS, Gonçalves RF, G. V. S. M. A.: 2021, Benford's law based method for fraud detection using R library, *MethodsX* (8).

Baum, C. F., Schaffer, M. E. and Stillman, S.: 2002, IVREG2: Stata module for extended instrumental variables/2SLS and GMM estimation, Statistical Software Components, Boston College Department of Economics.
**URL:** *https://ideas.repec.org/c/boc/bocode/s425401.html*

Dincecco, M., Fenske, J., Menon, A. and Mukherjee, S.: 2022, Pre-Colonial Warfare and Long-Run Development in India, *Economic Journal* **132**(643), 981–1010.
**URL:** *https://doi.org/10.1093/ej/ueab089*

Durtschi, C., Hillison, W. and Pacini, C.: 2004, The effective use of Benford's law to assist in detecting fraud in accounting data, *Journal of Forensic Accounting* **5**(1), 17–34.

Gaure, S.: 2022, *lfe: Linear Group Fixed Effects*.
**URL:** *https://CRAN.R-project.org/package=lfe*

Jaques, T. and Showalter, D.: 2006, *Dictionary of Battles and Sieges: A Guide to 8,500 Battles from Antiquity Through the Twenty-first Century*, Greenwood Press.
**URL:** *https://books.google.co.uk/books?id=Dh6jydKXikoC*

Kalyvas, S. N.: 2008, *Promises and pitfalls of an emerging research program: the microdynamics of civil war*, Cambridge University Press, p. 397–421.

Miller, C. and Bakar, K. S.: 2023, Conflict events worldwide since 1468BC: Introducing the historical conflict event dataset, *Journal of Conflict Resolution* **67**(2-3), 522–554.
**URL:** *https://doi.org/10.1177/00220027221119085*

Pun-Cheng, L. S.: 2016, *Distance Decay*, John Wiley  Sons, Ltd, pp. 1–5.
**URL:** *https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118786352.wbieg0179*

Python Core Team: 2015, *Python: A dynamic, open source programming language*, Python Software Foundation.
**URL:** *https://www.python.org/*

R Core Team: 2021, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
**URL:** *https://www.R-project.org/*

Raleigh, C., Linke, A., Hegre, H. and Karlsen, J.: 2010, Introducing ACLED: An armed conflict location and event dataset: Special data feature, *Journal of Peace Research* **47**(5), 651–660.

Sauerbrei, W., Meier-Hirmer, C., Benner, A. and Royston, P.: 2006, Multivariable regression model building by using fractional polynomials: description of SAS, STATA and R programs, *Computational statistics & data analysis* **50**(12), 3464–3485.

Sundberg, R. and Melander, E.: 2013, Introducing the UCDP georeferenced event dataset, *Journal of Peace Research* **50**(4), 523–532.

Vincenty, T.: 1975, Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations, *Survey review* **23**(176), 88–93.

Zeileis, A. and Hothorn, T.: 2002, Diagnostic checking in regression relationships, *R News* **2**(3), 7–10.
**URL:** *https://CRAN.R-project.org/doc/Rnews/*