

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

Rachel Forshaw 

Department of Accounting, Economics and Finance, Heriot-Watt University

Tim Ölkers

*Department of Agricultural Economics and Rural Development,
Georg-August-Universität Göttingen*

Ritika Sethi

Department of Economics, Rice University

Manali Sovani

Department of Economics, Tufts University

Abstract. We test the reproducibility and replicability of M. Dincecco, J. Fenske, A. Menon and S. Mukherjee (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 using the provided replication package in Stata, but note that the ease of replication could be improved by the provision of code and intermediate data sets for the conflict exposure measure. We test for and find no evidence of data manipulation in the final data sets. Concerning direct replicability, we consider different ways of measuring distance to conflicts and also alternative proxies for both the dependent variable and variables that capture channels by which the main effects operate. We find that some estimates are sensitive to the type of conflict considered. Other estimates are sensitive to the time period considered, most likely due to time heterogeneity in the number of conflicts recorded. Nevertheless, most estimates are substantially in line with the original study.

Résumé. Commentaire sur Dincecco, Fenske, Menon et Mukherjee (2022) : guerres précoloniales et développement à long terme en Inde. Nous testons la reproductibilité et la répétabilité de Dincecco, Fenske, Menon et Mukherjee, qui fait état d'une relation positive entre les guerres interétatiques précoloniales et les schémas de développement à long terme en l'Inde. Nous confirmons que tous les résultats de l'étude sont informatiquement reproductibles à l'aide du jeu de données de reproductibilité dans le logiciel Stata, mais notons que la reproductibilité pourrait être facilitée si on fournissait le code et les jeux de données intermédiaires pour la mesure de l'exposition au conflit. Nos tests ne révèlent aucune manipulation dans les jeux de données finaux. Aux fins de répétabilité directe, nous envisageons

Corresponding author: Rachel Forshaw, r.forshaw@hw.ac.uk

This replication study began as part of the Calgary Replication Games organized by the Institute for Replication (<https://i4replication.org/>). We thank Abel Brodeur and participants of the Calgary Replication Games as well as Mark Dincecco, James Fenske, Anil Menon, Shivaji Mukherjee and two anonymous referees for their comments and suggestions. Any remaining errors are our own.

Canadian Journal of Economics / *Revue canadienne d'économie* 2023 0(0)

xxxx 2023. / xxxx 2023.

0 / pp. 1–22 / DOI: 10.1111/caje.12693

© The Authors. Canadian Journal of Economics/Revue canadienne d'économie published by Wiley Periodicals LLC on behalf of Canadian Economics Association.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

différentes façons de mesurer la distance aux conflits, ainsi que d'autres substituts pour la variable dépendante et les variables qui saisissent les voies qui véhiculent les principaux effets. Nous constatons que certains résultats sont sensibles au type de conflit considéré. D'autres sont sensibles à la période considérée, très probablement en raison de l'hétérogénéité du nombre de conflits enregistrés au fil du temps. Néanmoins, la plupart des résultats sont en grande partie conformes à l'étude d'origine.

JEL classification: H11, N45, P48

1. Introduction

DINCECCO ET AL. (2022) 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 250 km 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 that 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 all of the results of Dincecco et al. (2022) 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. 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 Dincecco's et al. (2022) 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 (60.42% of results replicated) and time period over which the conflict exposure is considered (66.67% of results replicated). We also explore alternative ways of measuring conflict exposure using data from the Historical Conflict Event Dataset (Miller and Bakar 2023). Here, we are able to replicate the sign and significance of estimates in 100% of alternative measures of conflict exposure. 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 Dincecco's et al. (2022) main results and channels by which their main effects operate.

Finally, we re-examine one of the channels by which Dincecco's et al. (2022) 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 using new data from the Uppsala Conflict Data Program (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 organized 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 show reproducibility of the main results using the Stata code provided by Dincecco et al. (2022) in section 2.1. Secondly, in section 2.2, we test for data manipulation in the data sets provided in Dincecco et al. (2022) because only final data sets were provided by the authors and we find no evidence of data manipulation.

2.1. Stata reproducibility

Both the code and full data sets are provided by the authors and published on the Economic Journal website, available at <https://doi.org/10.5281/zenodo.5583263>. 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 Dincecco et al. (2022), the OLS specification is

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

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, μ_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). 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 because the Khyber Pass was the main route for invaders coming from Central Asia to India. We found no discrepancies between the original paper and the Stata reproductions.

2.2. Distribution of first digits

Dincecco et al. (2022) provide the final data sets, but not the base data sets nor the code to create them. As such, it was not possible to check data definitions or to recode key variables. Instead, we compare the relative frequency distribution of leading digits in each data set against Benford's law, which states that the proportion p_n of observations beginning with first significant digits $n = 1, 2, \dots, 9$ is approximately $\log_{10}(n+1)$. This law will hold where data are scale-invariant, i.e., independent of the units chosen.¹ Comparing data with this theoretical distribution is a technique used to look for fraud and manipulation in financial records and other data sets in forensic accounting. This test is most likely to capture

1 See Pinkham (1961) for a formal proof.

TABLE 1

Pre-colonial conflict and economic development: Main results

Dependent variable	Original study		
	ln(0.01 + Luminosity)		
	(1)	(2)	(3)
Pre-colonial conflict exposure	3.713*** (0.305)	1.601*** (0.380)	1.465*** (0.370)
Population density	Yes	Yes	Yes
State FE	No	Yes	Yes
Geographic controls	No	No	Yes
R^2	0.598	0.829	0.849
Observations	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 are 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% levels.

TABLE 2

Pre-colonial conflict and economic development: IV

Panel A: First stage Dependent variable:	Original study		
	Pre-colonial conflict exposure		
	(1)	(2)	(3)
Proximity to Khyber Pass	0.204*** (0.018)	0.094*** (0.025)	0.080*** (0.024)
Population density	Yes	Yes	Yes
State FE	No	Yes	Yes
Geographic controls	No	No	Yes
R^2	0.415	0.645	0.665
Observations	660	660	660
Panel B: Second stage Dependent variable:	ln(0.01 + Luminosity)		
	(1)	(2)	(3)
Pre-colonial conflict exposure	4.930*** (0.609)	4.626*** (1.291)	3.482** (1.389)
Population density	Yes	Yes	Yes
State FE	No	Yes	Yes
Geographic controls	No	No	Yes
R^2	0.593	0.814	0.843
Observations	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 and 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 are 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% levels.

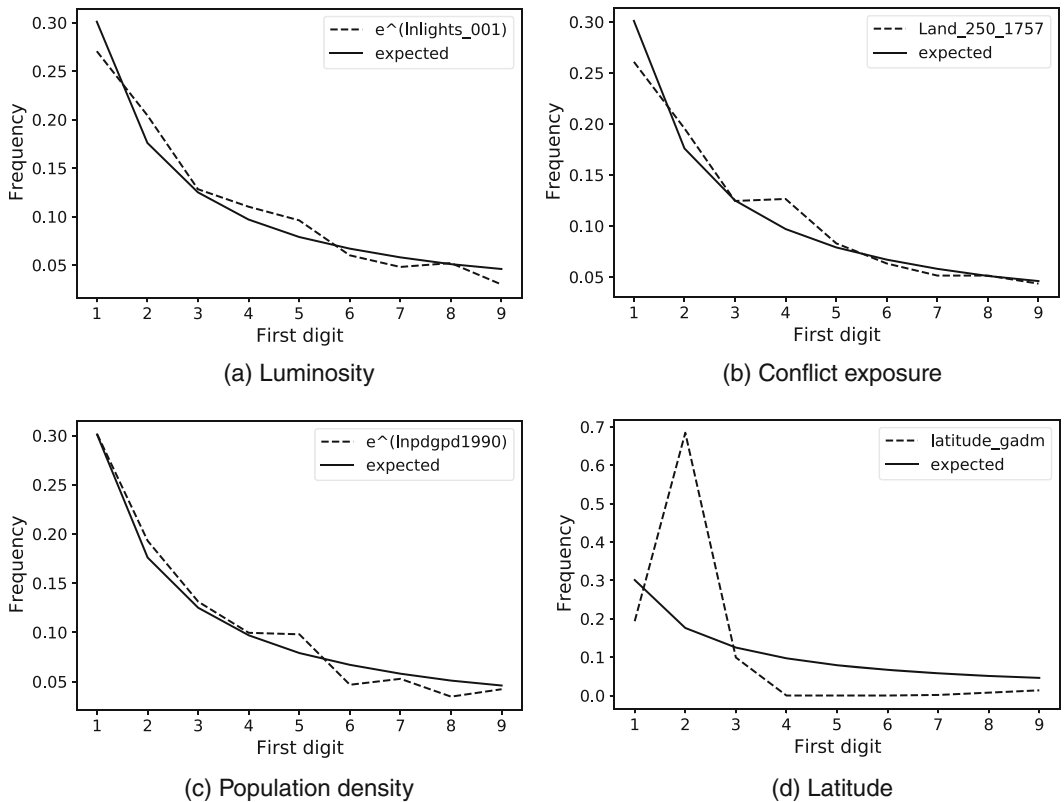


FIGURE 1 Empirical Benford distributions

NOTES: Panel A: Proxy for economic development as measured by luminosity the dependent variable, exponentiated to recover the distribution of digits. These data fit 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 250 km. These data fit 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. These data fit the expected distribution with a mean squared error (MSE) of 0.0002. Panel D: Latitude as measured using district centroids. These data fit the expected distribution with a mean squared error (MSE) of 0.03, which fails the Benford test, which we would expect given the assigned nature of the data.

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. We calculate distributions for scale-invariant variables in all 19 data sets provided by Dincecco et al. (2022) and compare these distributions to the expected distribution. Figure 1 shows examples of calculated distributions compared with the expected distributions for the variables used in the main specifications of Dincecco et al. (2022). Following da Silva Azevedo et al. (2021) we define a calculated distribution as not conforming to the expected distribution when the mean squared error (MSE) is in excess of 0.015.² Panels A to C, which show measures of Luminosity, Conflict Exposure and Population Density, all

² Other measures that are commonly used for anomaly detection include Chi-squared and Z-tests, which we do not consider because all data sets 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.

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. Because this is latitude data, we would not expect it to do so, given that the Indian subcontinent can lie only within given latitudes, i.e., this data is not scale-invariant. Repeating this process for all scale-invariant variables in the 19 data sets provided by Dincecco et al. (2022), we find no evidence of data manipulation.

3. Direct replicability

In this section, we test the ability to duplicate the results of Dincecco et al. (2022) 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 Dincecco et al. (2022) 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 Dincecco et al. (2022) but not reported. In section 3.3, we also consider the replicability of the channels analysis by Dincecco et al. (2022), which considers the predictions implied by their theoretical framework concerning pre-colonial conflict exposure and colonial state-making. 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).

3.1. Alternative proxies for conflict exposure

Dincecco et al. (2022) use land battles between 1000 and 1757 within 250 km of the district centroid as the main proxy for conflict exposure. The authors use historical conflict from Jaques (2006), which they geocode to create a 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 data set provided by Dincecco et al. (2022). These measures are *All conflicts*, *Multi-day*, *Multi-year*, *Naval*, *One-day*, *Sacking & razing*, *Siege* and *Storming battles*. Table 3 reports OLS and second stage IV estimates from the original study (row 1) and the estimates using alternative proxies (rows 2 to 9). 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* and *All conflicts* proxies replicate the results of Dincecco et al. (2022) closely because 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 (1) and (2) of table 3, respectively), encompassing *Multi-day*, *Multi-year*, *Naval* (of which the estimates are negative in the IV estimates), *Siege* in the OLS specification and *Storming* in both. Dincecco et al. (2022) argue that such alternative proxies were more likely to capture battles that affected the capital stock, diminishing the proposed mechanisms.

3.2. Alternative measures of conflict exposure

Dincecco et al. (2022) 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 in the online

TABLE 3

Pre-colonial conflict and economic development: Alternative conflict proxies

Dependent variable	ln(0.01 + Luminosity)	
	OLS (1)	IV (2)
Land battles (original study)	1.465*** (1.389)	3.482** (1.389)
All conflicts	0.681*** (0.250)	2.791** (1.094)
Multi-day	0.551 (0.496)	-69.692 (226.214)
Multi-year	-0.324 (3.462)	-98.369 (86.406)
Naval	-5.511 (6.326)	-1599.681 (3093.860)
One-day	1.256*** (0.358)	2.612*** (0.978)
Sacking, razing	4.796*** (1.697)	8.862*** (3.344)
Siege	0.320 (0.422)	13.943* (7.691)
Storming	-0.519 (0.708)	-156.537 (582.155)
Population density	Yes	Yes
State FE	Yes	Yes
Geographic controls	Yes	Yes
Observations	660	660

NOTES: Estimation method is OLS (column (1)) and 2SLS (column (2)) using data from Dincecco et al. (2022). Alternative proxies for conflict are provided by the authors in the original data set. 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. Significant at the ***1%, **5%, *10% levels.

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 alternative measures of exposure to conflict.³ We begin by replicating the measure used in the paper, and then consider several alternative measures.

Dincecco et al. (2022) 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:

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

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)

3 We thank an anonymous referee for this suggestion.

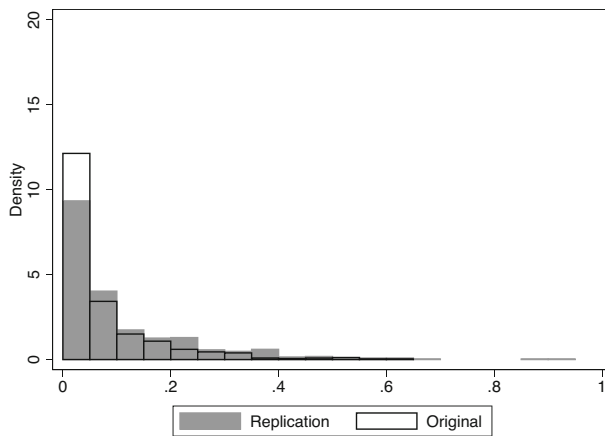


FIGURE 2 Conflict exposure measure distributions

NOTE: Distribution of the conflict exposure measure by district provided by Dincecco et al. (2022) (solid grey) versus replication using data from the Historical Conflict Event Dataset from Miller and Bakar (2023) (clear with black outline.)

and construct the measure of exposure to conflict in equation (2), keeping only land battles that occurred in the Indian subcontinent⁴ 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 2 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 6 compares the number of HCED conflicts with the number of conflicts calculated using data provided by Dincecco et al. (2022).⁵ There are fewer conflicts in total in the HCED data, concentrated between the periods 1000–1100 and post-1500.

In columns (1) and (2) of table 4, we compare how the results of the OLS and IV specifications differ when we use the reconstructed measure (panel B) in place of that provided by (Dincecco et al. 2022, panel A). The main parameter of interest, i.e., the effect of exposure to conflict on present-day development, is still positive and statistically significant but at a marginally smaller magnitude. As discussed in the preceding paragraph, measures using HCED data do not exactly match those of Dincecco et al. (2022). Dincecco et al. (2022) 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 Dincecco et al. (2022) and report the estimates in panel B' of table 4. Because Dincecco et al. (2022) provide a crosswalk for only the measure *All conflicts*, the relevant comparison is row 2 of table 3. Here we see

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

5 Dincecco et al. (2022) 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.

TABLE 4

Pre-colonial conflict and economic development: Alternative measures of conflict exposure

Dependent variable	ln(0.01 + Luminosity)	
	OLS (1)	IV (2)
A: Land battles 1001–1757 (original study)	1.465*** (0.370)	3.482** (1.389)
Standardized beta coefficient	0.095	0.225
R^2	0.849	0.843
B: Land battles 1001–1757 (replication using HCED)	0.717** (0.279)	3.183** (1.272)
Standardized beta coefficient	0.063	0.278
R^2	0.847	0.831
B': Conflicts 1001–1757 (replication using Dincecco et al. 2022)	0.678*** (0.249)	2.773** (1.085)
Standardized beta coefficient	0.066	0.268
R^2	0.847	0.831
C: Conflicts 1001–1757 (units = 10 km)	0.136*** (0.036)	0.373*** (0.138)
Standardized beta coefficient	0.093	0.257
R^2	0.848	0.840
D: Conflicts 1001–1757 (units = 100 km)	0.035*** (0.008)	0.069*** (0.024)
Standardized beta coefficient	0.121	0.238
R^2	0.849	0.846
E: Land battles 1001–1757 (Gaussian)	0.046*** (0.011)	0.105*** (0.037)
Standardized beta coefficient	0.103	0.238
R^2	0.848	0.843
F: Land battles 1001–1757 (count)	0.016*** (0.003)	0.031*** (0.011)
Standardized beta coefficient	0.126	0.240
R^2	0.849	0.846
Population density	Yes	Yes
State FE	Yes	Yes
Geographic controls	Yes	Yes
Observations	660	660

NOTES: Estimation method is OLS (column (1)) and 2SLS (column (2)). Variable of interest is pre-colonial exposure to land battles between 1000 and 1757 within 250 km of the district centroid, with alternative measurement as follows. A: Estimates reported by Dincecco et al. (2022). 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 using units of 10 km, i.e., dividing $distance_{i,c}$ by 10 before taking the inverse. D: Distance measure calculated using units of 100 km, i.e., dividing $distance_{i,c}$ by 100 before taking the inverse. E: Gaussian distance measure. F: Count of land battles within 250 km. 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% levels.

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. Using HCED data, we now summarize the alternative measures of distance to conflict that we consider.

In equation (2), Dincecco et al. (2022) add one to $\text{distance}_{i,j}$ before taking the inverse because not including 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.⁶ Nevertheless, this measure still gives a lot of weight to nearby conflicts and is not unit-invariant in relative terms. In order to explore the sensitivity of estimates to variation induced by smaller distances, we first explore how the results may change if we scale distances by factor $u > 0$, i.e.,

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

As u increases, the weight put on a given conflict decays less rapidly as its distance from the centroid increases i.e., conflicts that are further away contribute more to conflict exposure⁷. In the first and second alternative measures, we use the baseline measure as in equation (2) but explore using different distance units: 10 km and 100 km. If originally a distance from conflict a to district centroid b measured x kilometres, we now measure $x/10$ kilometres and $x/100$ kilometres, respectively. Here, we still follow their benchmark of including conflicts that occurred between 1000 and 1757, and within 250 km of the district. Comparing the the OLS and IV specifications using these measures, panels C and D compared with panel B of table 4 show that estimates are decreasing in magnitude in both the OLS and IV specifications, respectively, as u increases, however, the standardized coefficients are broadly unchanged and the sign and significance of the estimates are unaffected.

Secondly, we explore a Gaussian transformation of $\text{distance}_{i,c}$ (measured in 100 kilometres):

$$\sum_{c \in C} \exp(-\text{distance}_{i,c}), \quad (4)$$

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 Dincecco et al. (2022). Panel E of table 4 shows that the sign and significance of the estimated coefficients are unaffected, and the magnitude of the standardized coefficients are marginally higher.

Finally, we define a simple count-based measure of exposure to conflict, which is calculated as the number of conflicts within 250 km of the district centroid. Note that this corresponds to the the limiting case in equation (3) as $u \rightarrow \infty$. Panel G of table 4 shows that the sign and significance of the estimated coefficients are again unaffected, and magnitudes are marginally higher.

In conclusion, 100% of tests using alternative measures of conflict exposure replicate the sign, significance and are close in magnitude to the original paper's results.

6 The authors also show that the results are also robust to simply using the inverse of distance.

7 Re-expressing equation (3) as $\sum_{c \in C} (k + \text{distance}_{i,c}/u)^{-\epsilon}$, we note that the same effect could be achieved by increasing the k term or by decreasing the absolute magnitude of the elasticity parameter, ϵ .

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 Dincecco's et al. (2022) data. In this section, we therefore test the sensitivity of Dincecco's et al. (2022) results to the period of time over which pre-colonial conflict exposure is measured. In table 5, we replicate the main OLS results (table 1 of Dincecco et al. 2022) and IV specification (table 2 of Dincecco et al. 2022), respectively, but break down the pre-colonial conflict exposure variable into 100 year periods. Concretely, row 1 is exposure during years 1000–1757 (ie, replicating the original study). Row 2 reports the estimates for the time period 1000–1100, row 3 for exposure during years 1101–1200 and so on until row 8 for years 1601–1700. Note that for the time period 1400–1500, estimates are omitted because there were no recorded conflicts in the Dincecco et al. (2022) 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 5, 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, and in this specification, time block 1601–1700 now shows an insignificant effect. We note that Dincecco et al. (2022) 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 6 in our paper reports the number of pre-colonial conflicts, while table 7 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. However, 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. In order to analyze this further, we are able to examine the time heterogeneity in channels by which the main effect operates according to Dincecco's et al. (2022) theoretical framework. We therefore use this sub-period measure to estimate the remaining results of the paper, i.e., their tables 3 to 9.

Firstly, table 8 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 Dincecco et al. (2022) 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 8, 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 9 replicates the regression of colonial fiscal development on pre-colonial conflict exposure. Dincecco et al. (2022) report a positive and significant relationship between 1000–1757, which the authors argue is suggestive evidence that pre-colonial conflict exposure played a role in colonial-era state-making. Using the later time period for conflict exposure measure provided in the original study, 1500–1700, in row 2, we show results that are broadly consistent with the measures examined in the original study. 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

TABLE 5
Pre-colonial conflict and economic development: Main results using different time periods

	ln(0.01 + Luminosity)															
	OLS								IV							
Pre-colonial conflict exposure	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	
1000–1757 (original study)	1.465*** (0.370)							3.482*** (1.441)								
1000–1100 (replication)		11.05*** (3.732)							21.06*** (8.216)							
1101–1200 (replication)			7.779*** (2.665)							25.73*** (12.42)						
1201–1300 (replication)				5.502* (3.100)							14.25*** (6.207)					
1301–1400 (replication)					4.206* (2.430)							60.53* (32.91)				
1401–1500 (replication)																
1501–1600 (replication)						2.027* (1.039)								12.34*** (5.395)		
1601–1700 (replication)							3.040*** (0.860)								-45.57 (54.26)	
Population density	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Geographic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Observations	660	660	660	660	660	660	660	660	660	660	660	660	660	660	660	
R-squared	0.847	0.847	0.847	0.847	0.846	0.846	0.847	0.849	0.846	0.838	0.843	0.731	0.835	0.376		

NOTES: Estimation method is OLS (columns 1) to (9) and 2SLS (columns 9) to (16) 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 occurred 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% levels.

TABLE 6

Count of conflicts by time period

Century	N, using Dincecco et al. (2022)	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 Dincecco et al. 2022) and the Historical Conflict Event Dataset (HCED, Miller and Bakar 2023). Calculations by the present authors.

TABLE 7

Descriptive statistics of conflict exposure by time period

Century	N	Mean	Standard deviation	Min.	Max.
1000–1100	662	0.003	0.007	0	0.11
1101–1200	662	0.004	0.001	0	0.15
1201–1300	662	0.004	0.014	0	0.25
1301–1400	662	0.003	0.011	0	0.20
1401–1500	662	0	0	0	0
1501–1600	662	0.012	0.02	0	0.25
1601–1700	662	0.016	0.027	0	0.44
1701–1757	662	0.06	0.07	0	0.41
Total 1000–1757	662	0.07	0.01	0	0.61

NOTES: Mean of conflict exposure to land battles from 1001 to 1757 within 250 km of district centroid calculated over districts by century using data from Dincecco et al. (2022). Calculations by the present authors.

and by person. Again, we find that estimated magnitudes are marginally greater during the later years, 1500–1757.

Table 10 replicates table 5 of Dincecco et al. (2022), which examines the relationship between pre-colonial conflict, colonial and post-colonial conflict. 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. 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 1 of table 11 reports the replication of table 6 of Dincecco et al. (2022), which estimates a negative relationship between pre-colonial conflict and post-colonial political violence. The authors argue that reductions in political violence should result from greater colonial state-making. Row 2 shows the replication with the subset of conflicts, finding similar results. While the relationship with linguistic fractionalization (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) because this data set does not include the later time period data for pre-colonial conflict, and we are able to match

TABLE 8

Pre-colonial conflict and pre-colonial-era state-making

Dependent variable:	Important Mughal sites	State history		
		Babur (2)	Akbar (3)	Aurangzeb (4)
Pre-colonial conflict exposure	(1)			
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** (0.590)	0.552*** (0.162)	0.292 (0.206)	0.719*** (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) to (4) are state longevity in terms of districts incorporated into the Mughal Empire by Babur (1526–1530), 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 $\ln(\text{PopulationDensity})$ in 1500. Robust SEs in parenthesis are calculated using the robust command in Stata. Significant at the ***1%, **5%, *10% levels.

observations for only a subset of the original data.⁸ Nevertheless, we estimate a relationship that is similar to the original study.

Table 12 reports the results from replicating Dincecco’s et al. (2022) 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 Dincecco’s et al. (2022) results as the data containing irrigation rates and crop yields does not contain the conflict exposure data for 1500–1757. We are able to match data for only 208 of the original 271 observations. With this subset of data, the relationship with irrigation rates averaged between 1956 and 1987 (column (2)), and the relationship with crop yield

⁸ It is likely that it would be possible to recover all observations were a state and district crosswalk provided in the replication package.

TABLE 9
Pre-colonial conflict and colonial fiscal development

Dependent variable:	1881				1931			
	All		British India		Princely states		British India	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Pre-colonial conflict exposure	ln(Tax/Area)	ln(Tax/Person)	ln(Tax/Area)	ln(Tax/Person)	ln(Tax/Area)	ln(Tax/Person)	ln(Tax/Acre)	ln(Tax/Person)
1000–1757 (original study)	2.246*** (0.550)	1.245*** (0.382)	2.208*** (0.516)	1.157*** (0.354)	6.386* (3.524)	1.612 (2.228)	1.310 (0.911)	0.835 (0.705)
1500–1757 (replication)	2.335*** (0.659)	1.304*** (0.446)	2.173*** (0.589)	1.119*** (0.396)	10.41*** (3.766)	3.391 (2.585)	1.298 (1.120)	0.757 (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) to (6) and 1930 in columns (7) and (8). Robust SEs in parenthesis are calculated using the robust command in Stata. Significant at the ***1%, **5%, *10% levels.

TABLE 10

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	
	Land battles (1)	All conflicts (2)	Land battles (3)	All conflicts (4)	Land battles (5)	All conflicts (6)
Pre-colonial conflict exposure						
1000–1757 (original study)	0.170*** (0.036)	0.446*** (0.090)	0.037 (0.039)	0.313 (0.306)	−0.024*** (0.005)	−0.030*** (0.007)
1500–1757 (replication)	0.201*** (0.0497)	0.565*** (0.122)	0.101** (0.0497)	0.618 (0.469)	−0.0250*** (0.00632)	−0.0288*** (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) and (4), and post-colonial conflict exposure between 1947 and 2010 in columns (5) and (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 $\ln(\text{PopulationDensity})$ in 1750 in columns (1) and (2), in 1850 in columns (3) and (4) and in 1950 in columns (5) and (6). Robust SEs in parenthesis are calculated using the robust command in Stata. Significant at the ***1%, **5%, *10% levels.

TABLE 11

Pre-colonial conflict and post-colonial political violence

Dependent variable:	Political violence (1)	Maoist control (2)	Fractionalization	
			Linguistic (3)	Religious (4)
Pre-colonial conflict exposure				
1000–1757 (original study)	−0.241** (0.102)	−0.381** (0.163)	−0.209* (0.113)	0.080 (0.071)
R^2	0.408	0.281	0.570	0.557
Observations	660	395	660	660
1500–1757 (replication)	−0.329** (0.133)	−0.276* (0.162)	−0.210 (0.148)	0.0258 (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 Maoist control, a dummy variable that equals 1 for Maoist control in 2003. Dependent variable in column (3) is Linguistic fractionalization, defined as 1 minus the Herfindahl index of language population shares in 2001. Dependent variable in column (4) 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 $\ln(\text{PopulationDensity})$ in 1990. Robust SEs in parenthesis are calculated using the robust command in Stata. Significant at the ***1%, **5%, *10% levels.

TABLE 12

Pre-colonial conflict and irrigation infrastructure

Dependent variable:	% Irrigated		ln(Yield)	% Non-agriculture
	1931	1956–1987		
	(1)	(2)	(3)	(4)
Pre-colonial conflict exposure				
1000–1757 (original study)	21.275** (10.357)	37.413** (15.758)	0.737* (0.381)	0.197** (0.085)
R^2	0.391	0.611	0.683	0.566
Observations	257	271	271	660
1500–1757 (replication)	15.81 (10.47)	39.40 (24.95)	0.168 (0.621)	0.249** (0.107)
Constant	17.27 (29.93)	–23.88 (65.80)	0.860 (1.725)	0.416 (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) and (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% levels.

(column (3)) is no longer significantly different from zero. However, it impossible to ascertain whether this is due to the later time period considered for conflict exposure, or due to missing observations.

Table 13 replicates table 8 in Dincecco et al. (2022), which estimates the relationship between pre-colonial conflict exposure and district-level literacy rates under British colonial rule. Dincecco et al. (2022) estimate no relationship for literacy rates in 1881 and 1921 but strong positive relationships in 1961–1991 and 2011 (upper panel of table 13). We are able to directly replicate the results in columns 1 and 2 using the conflict exposure data from 1500 to 1757 provided by the authors, and also find no significant relationships. For literacy rates in 1961–1991 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.

Finally, table 14 re-examines the relationship between pre-colonial conflict exposure and education provision by district. In column (1), the variable of interest is the proportion of villages with primary schools in 1981, column (2) measures the proportion with high

TABLE 13

Pre-colonial conflict and literacy

Dependent variable:	% Literacy			
	1881 (1)	1921 (2)	1961–1991 (3)	2011 (4)
1000–1757 (original study)	−1.933 (3.188)	−5.635 (3.772)	11.796* (6.888)	10.146** (4.119)
R^2	0.464	0.556	0.623	0.599
Observations	251	303	271	626
1500–1757 (replication)	−0.199 (3.529)	−4.827 (5.129)	12.19 (8.756)	1.615 (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 1961–1991 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 and above. 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 $\ln(\text{PopulationDensity})$ 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% levels.

TABLE 14

Pre-colonial conflict and literacy

Dependent variable:	% Primary (1)	% High (2)	% Infant mortality (3)
Pre-colonial conflict exposure			
1000–1757 (original study)	18.683* (11.150)	−16.094** (6.553)	−35.283** (14.405)
R^2	0.712	0.840	0.674
Observations	203	187	270
1500–1757 (replication)	32.982** (13.316)	−6.491 (4.163)	−33.871 (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 % Infant mortality 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 $\ln(\text{PopulationDensity})$ 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% levels.

schools and column (3) is the infant mortality rate in 1991. Dincecco et al. (2022) find a positive relationship with primary education provision, and negative relationships with high school provision and infant mortality rates. They interpret this finding as conflict exposure promoting state-making which favoured investment in basic human capital at the expense of advanced human capital investments. We estimate a relationship of the same significance for the Mughal empire period only for primary provision; the coefficients on high school and infant mortality are not significantly different from zero. We also note that the magnitude of the coefficient on primary education is much higher in the later time period. However, as again we are unable to match all observations, we cannot comment whether the differences are due to reduced observations or the effects of time period heterogeneity.

To further understand the importance of time heterogeneity in the number of conflicts recorded by century discussed in 3.2, this section replicated the results of Dincecco et al. (2022) over 100 year time blocks, finding heterogeneous time effects. In particular, restricting the sample to the periods 1501–1600 and 1601–1700, the time of the Mughal empire, resulted in estimates of a much smaller magnitudes. Re-examining the mechanisms via which the authors effects operate for the period 1500–1757, we find estimates of the same sign and significance during this time period in 66.67% of replications. Furthermore, estimated magnitudes for estimates of channels are in general higher for this time period, providing suggestive evidence that the heterogeneous time effects estimated in table 5 are more likely due to size effects than diminishing returns to conflict exposure.

3.4. Alternative political violence data source

One of the three predictions of Dincecco's et al. (2022) conceptual framework is that districts that were more exposed to pre-colonial conflict would experience lower political violence in the long-term as a result of greater state-making. In this section, we use alternate conflict data provided by the Uppsala Conflict Data Program (UCDP) (Sundberg and Melander 2013) to re-estimate the relationship between organized violence per district between 2015 and 2018, and pre-colonial conflict exposure. The UCDP data covers individual events of organized violence, which they define as the phenomena of lethal violence occurring at a given time and place.

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 Dincecco et al. (2022). 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 columns (2) and (3) of table 15. Using the UCDP data, we no longer find a negative and significant relationship between pre-colonial conflict exposure and contemporary political violence that happened using the same time period (between 2015 and 2018) and geographical level (district) as the Dincecco et al. (2022) paper. If we include conflicts between 2001 and 2021 (column (3), table 15), we do confirm the negative and significant relationship but note that it is sensitive to changes in time period considered.

4. Conclusion

Table 16 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 (table 16, 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 data sets provided by the authors (first digits). For direct replicability, we consider

TABLE 15

Pre-colonial conflict and post colonial political violence

	Original study	Replication using different data	
	(1) 2015–2018	(2) 2001–2021	(3) 2015–2018
Time period	2015–2018	2001–2021	2015–2018
Pre-colonial conflict exposure	−0.241** (0.102)	−0.639* (0.349)	−0.034 (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. Column (1) replicates the results. The dependent variable is Political violence, defined as fatalities per district between 2015 and 2018 (in hundreds) using ACLED data. The dependent variable in columns (2) and (3) is Organized violence, defined as fatalities per state (in hundreds) based on the UCDP data. Column (2) focuses 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 $\ln(\text{PopulationDensity})$ in 1990. Robust SEs in parenthesis are calculated using the robust command in Stata. Significant at the ***1%, **5%, *10% levels.

alternative measures of conflict exposure using the Historical Conflict Event Dataset (Miller and Bakar 2023). Because Dincecco et al. (2022) 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. We also examine alternative proxies for conflict exposure provided by the authors. Dincecco et al. (2022) argue that such alternative proxies were more likely to capture battles that affected the capital stock, diminishing the proposed mechanisms. Nevertheless, in 60.42% of tests (total, alternative conflict proxies) we are able to replicate the sign and significance reported by Dincecco et al. (2022), but note that some magnitudes of estimates also differ from the original paper. When considering different time periods between years 1000 and 1757 for the conflict exposure proxy, we are able to replicate results in 91.67% 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, another explanation could be 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 66.67% 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 Dincecco et al. (2022). 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 data sets 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

TABLE 16

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, table 3	All conflicts	6	6	100	
		Multi-day	6	2	33.33	
		Multi-year	6	2	33.33	
		Naval	6	1	16.66	
		One-day	6	6	100	
		Sacking	6	6	100	
		Siege	6	5	83.33	
		Storming	6	1	16.66	
	Total, alternative conflict proxies			48	29	60.42
	Alternative measures of conflict exposure, table 4	Replication	2	2	100	
		Distance unit: 10 km	2	2	100	
		Distance unit: 100 km	2	2	100	
		Gaussian	2	2	100	
		Count	2	2	100	
	Total, alternative conflict exposure measures			10	10	100
	Different time periods, table 5	1000–1100	2	2	100	
		1101–1200	2	2	100	
		1201–1300	2	2	100	
		1301–1400	2	2	100	
		1501–1600	2	2	100	
		1601–1700	2	1	50	
	Total, different time periods (main results)			12	11	91.66
	Different time periods	Table 8	4	2	50	
Table 9		8	8	100		
Table 10		6	5	83.33		
Table 11		4	3	75		
Table 12		4	1	25		
Table 13		4	2	50		
Table 14		3	1	33.33		
Total		33	22	66.66		
Total, different time periods (mechanisms results)			33	22	66.66	
Alternative political violence data	Table 15	2	1	50		
	Total	2	1	50		
Total direct			105	73	69.52	
Grand total			130	98	75.38	

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.

economic development than diminishing returns to exposure. Using an alternative proxy for political violence from the UCDP (Sundberg and Melander 2013), we are not able to replicate estimates using the time period for the dependent variable used in Dincecco et al. (2022) but are able to replicate the findings using a larger window including more recent data. Taken together, we confirm direct replicability in 69.52% of direct tests (total direct). Taking all results together (grand total), we find that 75.38% 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 data sets 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 substantially in line with the original study.

Supporting information

The data and code that support the findings of this study are available in the Canadian Journal of Economics Dataverse at <https://doi.org/10.5683/SP3/RYYYIQ>.

References

- da Silva Azevedo, C., R. F. Gonçalves, V. L. Gava, and M. de Mesquita Spinola (2021) “A Benford’s law based method for fraud detection using R library,” *MethodsX* 8, 101575
- Dincecco, M., J. Fenske, A. Menon, and S. Mukherjee (2022) “Pre-colonial warfare and long-run development in India,” *Economic Journal* 132(643), 981–1010
- Durtschi, C., W. Hillison, and C. Pacini (2004) “The effective use of Benford’s law to assist in detecting fraud in accounting data,” *Journal of Forensic Accounting* 5(1), 17–34
- Jaques, T., (2006) *Dictionary of Battles and Sieges: A Guide to 8,500 Battles from Antiquity Through the Twenty-first Century*. Westport, CT: Greenwood Press
- Miller, C., and K. S. Bakar (2023) “Conflict events worldwide since 1468BC: Introducing the Historical Conflict Event Dataset,” *Journal of Conflict Resolution* 67(2-3), 522–54
- Pinkham, R.S. (1961) “On the distribution of first significant digits,” *Annals of Mathematical Statistics* 32(4), 1223–30
- Pun-Cheng, L.S.C. (2016) “Distance decay.” In D. Richardson, N. Castree, M. F. Goodchild, A. Kobayashi, W. Liu and R. A. Marston, eds., *International Encyclopedia of Geography: People, the Earth, Environment and Technology*, pp. 1–5. Wiley-Blackwell
- Sundberg, R., and E. Melander (2013) “Introducing the UCDP Georeferenced Event Dataset,” *Journal of Peace Research* 50(4), 523–32
- Vincenty, T. (1975) “Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations,” *Survey Review* 23(176), 88–93