

## Reflection summary

Wenjing Ding, Jiechen Qiu, Sizhe Jia, Zexi Xu

This project uses publicly available Colonial Origins data to replicate the main OLS results from Acemoglu, Johnson, and Robinson (2001). They find that higher institutional quality is strongly linked to higher income levels. As an extension, we also created scatter plots for Africa, Asia, and other former colonies. In all regions, we observe a positive relationship between institutional quality and GDP per capita. This supports the idea that institutions are a key and universal driver of long-run economic development.

According to the guideline, we reproduced AJR's regressions which is though the settler mortality and finally reflect economic performance. The results match the original paper closely. The policy has a large and significant causal impact on development. In addition, the IV estimates are larger than the OLS estimates, which is expected because measurement error usually makes OLS underestimate the true effect. The regressions remain strong across different samples, supporting AJR's argument that settler mortality is a valid and powerful instrument.

Data cleaning is the process of detecting, correcting, or removing errors and inconsistencies in datasets to improve data quality. AJR argue that their results remain robust under reasonable data selection, such as trimming extreme epidemic mortality values and using multiple reliable historical sources like unjustified deletion of many countries, misclassification of campaign data, and an overly small sample driven by outliers.

We also find clear evidence of historical persistence in institutions. Early measures such as constraints on the executive around 1900, early democracy, and the share of European settlers. At the same time, it is useful to continue to explain modern institutional quality. These results are consistent with the argument made in AJR's 2012 Reply and further strengthen the *colonial origins hypothesis*.

Overall, this replication project helped us better understand identification strategies in econometrics, how to evaluate the validity of an instrumental variable, and the importance of robustness checks in empirical research. By replicating and extending AJR's work, we gained valuable experience in developing data and owing a deeper understanding of the methodological logic.