

Overview

This replication package accompanies the paper “Female Leaders and Intrahousehold Dynamics: Evidence from State Elections in India” by S Anukriti, Bilge Erten, and Priya Mukherjee. The package contains Stata code and supporting documentation required to reproduce all tables and figures reported in the main paper and online appendix. The analysis is conducted using Stata (version 17 for the original analysis). Replication proceeds by setting the directory path in MASTER.do and executing the file, which sequentially cleans raw data, constructs analysis datasets, and generates outputs.

Data Availability and Provenance Statements

This paper uses multiple secondary data sources from India. The authors are secondary users of these data. Access conditions differ by source, as described below.

Details on Each Data Source

1. District-Level Household and Facility Survey (DLHS)

Main analysis uses DLHS-4 (2012–2014), conducted by the Ministry of Health and Family Welfare, Government of India. Robustness checks use DLHS-2 (2002–2004) and DLHS-3 (2007–2008). Data access requires submitting a request to the International Institute for Population Sciences (IIPS). See <https://www.iipsindia.ac.in/content/data-center>.

2. National Family Health Survey (NFHS)

The analysis uses NFHS-4 (2015–2016), a nationally representative survey of women aged 15–49. Data are accessed through the DHS Program (<https://dhsprogram.com/>) with user registration.

3. Election Data

State legislative election data for 2008–2015 were obtained from the Election Commission of India (ECI). Data for 1996–2007 and 2015–2023 were sourced from Ashoka University’s Lok Dhaba dataset (<https://lokhaba.ashoka.edu.in/>). Original ECI spreadsheets, URLs, and webpage screenshots are provided in the replication folder.

4. Census Data

District characteristics are obtained from the 2001 and 2011 Censuses of India. 2001 Census data are accessed via SHRUG (<https://www.devdata.org/shrug>). 2011 Census data are from https://censusindia.gov.in/2011census/population_enumeration.html.

5. Candidate Affidavit Data

Candidate characteristics are sourced from the Association for Democratic Reforms (ADR) via <https://www.myneta.info/> for elections between 2008 and 2015.

Dataset Summary Table

Dataset	Files	Source	Access	Provided in Package
DLHS-2/3/4	INDIA_VILLAGE.DTA; DLHS3VIND.dta; DLHS-4 village.dta	IIPS, Government of India	Restricted (Request Required)	No, but codebook has been provided
NFHS-4	IAIR74FL.DTA; IACR74FL.DTA; IAMR74FL.DTA	DHS Program	Restricted (Registration)	No, but codebook has been provided
Election Data	18 XLS spreadsheets and original source information (“Records”) in “ECI Spreadsheets” folder and; All States AE.csv	ECI / Lok Dhaba	Public*	Yes
Census Data	DL_census_data.xlsx	Government of India	Public	Yes
Candidate Affidavits	31 State-wise CSV files	Association for Democratic Reforms	Public*	Yes
SHRUG**	assembly_candidates_clean.dta	Development Data Lab (SHRUG)	Public	Yes
	con_shrug_2008_nl_wide.dta			
	con_shrug_2008.dta			
	pc01_pca_clean_pc01dist.dta			
	shrid_loc_names.dta			
	shrid_pc01dist_key.dta			
	shrug_pc01_district_key.dta			
	shrug_pc01.dta			

*These files are included in the replication package. However, users attempting to directly access Indian government websites may encounter geographic access restrictions and may need to be physically located in India for access.

**We use multiple datasets from the *Socioeconomic High-resolution Rural-Urban Geographic Platform (SHRUG)*, (<https://www.devdatalab.org/shrug>). The appropriate citations (Asher et al. (2019); Henderson, Storeygard and Weil (2011); Asher et al. (2021); Office of the Registrar General and Census Commissioner (2011); Jensenius and Verniers (2017); Prakash, Rockmore and Uppal (2019)) are provided here and in the manuscript.

Computational Requirements

Software Requirements:

Stata (version 17 used for original analysis; replication tested on Stata 18 MP).

Hardware and Runtime Information (World Bank reproducibility environment):

Operating System: Windows 11 Enterprise (23H2)

Processor: Intel Xeon Platinum 8562Y+ (dual)

Memory: 1.5 TB RAM
Approximate runtime: ~30 minutes.

Controlled Randomness

No stochastic simulation or randomized procedures are used in the analysis. Nonetheless, random seeds are set in the code to ensure exact reproducibility of all outputs, including regression discontinuity figures that may rely on internally randomized routines.

Description of Programs / Code

All Stata do-files are located in the ~\Do Files\ directory. The master script MASTER.do executes all steps in the correct order.

MASTER.do runs 26 do-files that:

- Append and clean election data
- Clean census data
- Construct DLHS and NFHS analysis datasets (main and robustness bandwidths)
- Generate all main-text and appendix tables and figures

MASTER.do runs the following do files in the right order to clean and organize the raw data and to reproduce the tables and figures reported in the paper.

- **Files 1-3 append and clean election data**
 1. *December 2013 missing district names.do*
 2. *Data append elections.do*
 3. *Data cleaning elections.do*
- **File 4 cleans the Census data**
 4. *Census cleaning.do*
- **Files 5-8 prepare DLHS datasets for the main analysis bandwidth 0.03, and robustness check bandwidths 0.02, 0.025, and 0.035**
 5. *DLHS4 election merge_3_0.do*
 6. *DLHS4 election merge_2_0.do*
 7. *DLHS4 election merge_2_5.do*
 8. *DLHS4 election merge_3_5.do*

- **Files 9-14 prepare NFHS datasets for the woman-level analysis at bandwidth 0.03, and robustness check bandwidths 0.02, 0.025, and 0.035. They also prepare the male dataset and the woman panel dataset**

9. *NFHS (female) election merge_3_0.do*

10. *NFHS (female) election merge_2_0.do*

11. *NFHS (female) election merge_2_5.do*

12. *NFHS (female) election merge_3_5.do*

13. *NFHS (male) election merge.do*

14. *NFHS mother panel.do*

- **File 15 conducts the final cleaning for all the main figures and tables in the main text.**

15. *Final cleaning.do*

- **File 16 provides the code for all the main figures and tables in the main text.**

16. *Main Figures and Tables.do*. Regression discontinuity graphs use the “rd” program defined in “rd_an.do file.

- **File 17 prepares additional candidate-level data for Appendix Tables A.4 and A.10**

17. *Data Prep for Tables A4 and A10.do*

- **Files 18-22 prepare the 3 rounds of DLHS data for Appendix Table A.18**

18. *elections_2003-2004.do*

19. *Data append elections 1996-2007.do*

20. *Data cleaning elections 1996-2007.do*

21. *Census Cleaning_2001.do*

22. *DLHS234 election merge.do*

- **File 23 prepares the data for Appendix Table A.19**
23. All election years 2007-2023.do
- **File 24 prepares the final data for all appendix tables and figures.**
24. Final cleaning appendix.do
- **Generate appendix tables and figures**
 - 25. Appendix Figures:* This do file generates all appendix figures. Regression discontinuity graphs use the “rd” program defined in “rd_an.do file.
 - 26. Appendix Tables:* This do file generates all appendix tables. These include Appendix Tables A.1-A.26.

Each do file directly outputs a .tex file for tables and a .pdf file for figures, which also reproduce the results used in the paper. Please see the log file that is generated for F-statistics for balance tables (Appendix tables 5, 6, and 11).

Reconstructing the Final Datasets

The final constructed datasets needed to generate the tables and figures in the paper are listed below for the reader’s convenience. These datasets are not included in the replication package - to generate them, run the do files listed later in this document.

- 1. DLHS_data_for_analysis.dta:* This is the cleaned version of the DLHS-4 dataset. To go from the raw version to the cleaned version, we exclude villages with missing information on covariate data and create all the variables used in analysis.
- 2. NFHS_data_for_analysis.dta:* This is the cleaned version of the NFHS-4 dataset for our main analysis involving ever-married women aged 15-49. To go from the raw version to the cleaned version, we exclude women with missing information on covariate data and create all the variables used in analysis.
- 3. NFHS_woman_panel.dta:* This is the reshaped version of the NFHS dataset to construct a retrospective woman-year panel spanning the time period between the last state election and the year of survey.

4. *men_data_for_analysis.dta*: This is the cleaned version of the men's survey from the NFHS-4 dataset. To go from the raw version to the cleaned version, we exclude (i) men with missing district information and (ii) men with missing information on covariate data.
5. *NFHS_data_figures.dta*: This file includes the NFHS-4 data merged with election data for illustrating the main results in regression discontinuity plots.
6. *NFHS_woman_panel_figures.dta*: This file includes the retrospective woman-year panel from the NFHS-4 data merged with election data for illustrating the results for birth spacing outcome in regression discontinuity plots.
7. *DLHS_data_figures.dta*: This file includes the DLHS-4 data merged with election data for illustrating the main results in regression discontinuity plots.
8. *elections.dta*: This is the election data covering 2008-2015, and it is used to illustrate the first stage.
9. *census2001.dta*: This file includes data on predetermined district characteristics from the 2001 Census data and elections for conducting the balance test.
10. *DLHS_appended_for_analysis.dta*: This is the cleaned and appended version of the DLHS-2, DLHS-3, and DLHS-4 datasets. To go from the raw version to the cleaned version, we exclude (i) villages with missing district information and (ii) villages with missing information on covariate data.
11. *DLHS_2_for_analysis.dta*: This dataset includes DLHS-2 survey data along with 2011-2015 election and the 2001 Census data. These election years correspond to the second round of elections that occurred in the states after DLHS-2 survey years.
12. *DLHS_2_restricted_for_analysis.dta*: This is the restricted version of the *DLHS_2_for_analysis* dataset, which excludes districts in which there were close male-female elections where the female candidate won in the first round 1 (2006-2010) of elections after DLHS-2 survey years.
13. *DLHS_3_for_analysis.dta*: This dataset includes DLHS-3 survey data along with 2014-2018 election and the 2001 Census data. These election years correspond to the second round of elections that occurred in the states after DLHS-3 survey years.
14. *DLHS_3_restricted_for_analysis.dta*: This is the restricted version of the *DLHS_3_for_analysis* dataset, which excludes districts in which there were close male-

female elections where the female candidate won in the first round 1 (2009-2013) of elections after DLHS-3 survey years.

15. *Candidates_dta*: Data on electoral candidates between 2008 and 2015 are from the Election Commission of India.
16. *Candidates_background.dta*: Data are from candidate affidavit data provided by the Association of Democratic Reform for the years between 2008 and 2015.
17. *norms_nfhs.dta*: This is a district level data set using variables (averaged at the district level) from the NFHS data.
18. *norms_dlhs.dta*: This is a district level data set using variables (averaged at the district level) from the DLHS data.
19. *DLHS_data_for_analysis_2.dta*, *DLHS_data_for_analysis_2_5.dta*, *DLHS_data_for_analysis_3_5.dta*: These data files include the DLHS data for analysis, but for bandwidths 2, 2.5, and 3.5 percentage points for margin of victory, respectively, and are used for the robustness checks in the Appendix.
20. *NFHS_data_for_analysis_2.dta*, *NFHS_data_for_analysis_2_5.dta*, *NFHS_data_for_analysis_3_5.dta*: These data files include the NFHS data for analysis, but for bandwidths 2, 2.5, and 3.5 percentage points for margin of victory, respectively, and are used for the robustness checks in the Appendix.

Mapping Between Outputs and Manuscript

Each do-file outputs LaTeX (.tex) tables and PDF figures that correspond exactly to the exhibits in the paper. A complete mapping of figures and tables to output filenames is provided below:

Exhibit in the manuscript	Corresponding output from replication package
Figure 1	figure1_combined.pdf
Table 1	dlhs_health_facility.tex
Table 2	dlhs_health_worker.tex
Table 3	dlhs_programs.tex
Table 4	fprh.tex spacing.tex

Table 5	ipv.tex
Table 6	alt_channels_woman.tex alt_channels_man.tex
APPENDIX FIGURES AND TABLES	
Appendix Figure 1	figA1_mccrary_stata.pdf
Appendix Figure 2	figA2_balanced_covariates.pdf
Appendix Figure 3	figA3.pdf
Appendix Table 1	Panel A: tableA1_dist_dlhs_elec.tex Panel B: tableA1_dist_dlhs.tex Panel C: tableA1_dlhs.tex
Appendix Table 2	Panel A: tableA2_female.tex Panel B: tableA2_panel.tex Panel C: tableA2_male.tex
Appendix Table 3	Panel A: tableA3_dist_elec.tex Panel B: tableA3_dist.tex
Appendix Table 4	Column 1: top_2.tex Column 2: all_candidates.tex
Appendix Table 5	bal_dhs.tex
Appendix Table 6	bal_dlhs.tex
Appendix Table 7	correlations_fem_frac_norms_nfhs.tex
Appendix Table 8	corrleations_hw_norms.tex
Appendix Table 9	candidates_balance_ttests.tex
Appendix Table 10	male_winner_vote_share_winner.tex male_winner_winner_incumb_party.tex male_winner_turnout.tex male_winner_prev_close.tex male_winner_prev_close_mf.tex male_winner_tot_fem_cand.tex
Appendix Table 11	bal_dhs_anyclose.tex
Appendix Table 12	reduced_form.tex
Appendix Table 13	fprh_urban.tex spacing_urban.tex
Appendix Table 14	psycho_sexual_ipv.tex
Appendix Table 15	ipv_urban.tex
Appendix Table 16	robust_dlhs_hf.tex robust_dlhs_hw.tex robust_dlhs_p.tex
Appendix Table 17	tableA17_robust_physical_rural.tex

Appendix Table 18	dlhs_health_facility_A18.tex
Appendix Table 19	A19_falsification_dlhs2_round2.tex A19_falsification_dlhs2_restricted_round2.tex A19_falsification_dlhs3_round2.tex A19_falsification_dlhs3_restricted_round2.tex
Appendix Table 20	dlhs_health_facility_collapsed.tex
Appendix Table 21	dlhs_health_worker_collapsed.tex
Appendix Table 22	dlhs_programs_collapsed.tex
Appendix Table 23	fprh_collapsed.tex spacing_collapsed.tex fprh_collapsed-couples.tex spacing_collapsed-couples.tex
Appendix Table 24	ipv_collapsed.tex ipv_collapsed_young.tex ipv_collapsed_older.tex ipv_collapsed-gias_gt0.tex ipv_collapsed-gias_le0.tex
Appendix Table 25	A25_discord.tex
Appendix Table 26	A26_public_goods.tex

Instructions to Replicators

1. Obtain access to all restricted datasets listed above.
2. Place raw data files in the appropriate subdirectories as indicated in the replication folder.
3. Open MASTER.do and set the root directory path.
4. Run MASTER.do in Stata to reproduce all tables and figures.

Notes

- To the best of the authors' knowledge, there is no time restriction in terms of how long the data used in this article will be preserved in the restricted-access location. The sources do not provide any expiration date for the data.
- We commit to preserving data and code for a period of no less than five years following publication of the manuscript, and to providing reasonable assistance to requests for clarification and replication.

References

Asher, Sam, Tobias Lunt, Ryu Matsuura, and Paul Novosad. 2019. “The Socioeconomic High-resolution Rural-Urban Geographic Dataset on India (SHRUG).” Working paper.

Asher, Sam, Tobias Lunt, Ryu Matsuura, and Paul Novosad. 2021. “Development Research at High Geographic Resolution: An Analysis of Night Lights, Firms, and Poverty in India using the SHRUG Open Data Platform.” *World Bank Economic Review*.

Henderson, J. Vernon, Adam Storeygard, and David N. Weil. 2011. “A Bright Idea for Measuring Economic Growth.” *American Economic Review*.

Jensenius, Francesca R, and Gilles Verniers. 2017. “Studying Indian politics with large-scale data: Indian election data 1961–today.” *Studies in Indian Politics*.

Office of the Registrar General, Office, and India Census Commissioner. 2011. “Census of India.” Accessed via SHRUG [https://www.devdatalab.org/shrug_download/, v2.1.pakora](https://www.devdatalab.org/shrug_download/v2.1.pakora).

Prakash, Nishith, Marc Rockmore, and Yogesh Uppal. 2019. “Do criminally accused politicians affect economic outcomes? Evidence from India.” *Journal of Development Economics*.