

README: replication instructions for “Intergenerational Persistence in Child Mortality”

Overview

This file contains replication instructions for the paper “Intergenerational Persistence in Child Mortality”.

Data Availability and Provenance Statements

- ☐ This paper does not involve analysis of external data (i.e., no data are used or the only data are generated by the authors via simulation in their code).

If box above is checked and if no simulated/synthetic data files are provided by the authors, please skip directly to the section on Computational Requirements. Otherwise, continue.

Statement about Rights

- ☒ I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.
- ☒ I certify that the author(s) of the manuscript have documented permission to redistribute/publish the data contained within this replication package. Appropriate permission are documented in the LICENSE.txt file.

Summary of Availability

- ☒ All data **are** publicly available.
- ☐ Some data **cannot be made** publicly available.
- ☐ **No data can be made** publicly available.

Details on each Data Source

This project uses publicly available secondary data from the following sources:

- Demographic and Health Surveys (ICF, 2004-2017)
- UCDP/PRIO Armed Conflict Dataset (version 20.1) (Gleditsch et al., 2002; Pettersson et al., 2021).
- Penn World Table (version 9.1) for GDP data (Feenstra et al., 2015).
- United Nation World Population Prospects data (2019 revision) for child mortality rates by country x time period (United Nations and Social Affairs, 2019).

More details on each data source below. The datasets for UCDP-PRIO, Penn World Table (version 9.1) and United Nation World Population Prospects data (2019 revision) are provided in the replication directory “data” folder. The DHS data must be obtained by the user directly (see instructions below).

Demographic and Health Surveys

Due to the DHS program’s Terms of Use, we are not able to provide a copy of the DHS data in our replication directory. The user must register as a data user on the DHS website and download the raw data. Instructions to access DHS data are here: <https://dhsprogram.com/data/Access-Instructions.cfm>.¹

See Section “Instructions to Replicators” for detailed instructions on how to download the DHS datasets required for replication.

Data files: N/A

Codebook: Codebooks for DHS Standard Recodes can be found at: <https://dhsprogram.com/publications/publication-dhsg4-dhs-questionnaires-and-manuals.cfm>

UCDP/PRIO

Data for conflict events from the UCDP/PRIO Armed Conflict Dataset version 20.1 were downloaded from the Uppsala Conflict Data Program website at: https://ucdp.uu.se/downloads/replication_data.html. We use version 20.1 of the dataset in our analysis and it is provided in the replication directory.

Note: The UCDP/PRIO has been updated to a new version since our project was completed. They maintain documentation of their complete version history here: https://ucdp.uu.se/downloads/replication_data.html. Very few modifications were made between version 20.1 and 21.1 (see section “Changes from Version 20.1 to 21.1”), so this update should not substantially impact our analysis results if at all.

Datafile: “data/aggregates/ucdp_prio/ucdp-prio-acd-201.dta” (.dta format)

Codebook: “data/aggregates/ucdp_prio/ucdp-prio-acd-201.pdf” (.pdf format)

Penn World Table (version 9.1)

Data on GDP from the Penn World Table were downloaded from: <https://www.rug.nl/ggdc/productivity/pwt/pwt-releases/pwt9.1> (DOI: 10.15141/S50T0R). A copy of the data is provided in the replication directory.

Note: The PWT has been updated to a new version since our project was completed. There were a small number of minor changes, so this update should not substantially impact our analysis results. Please refer to the change log of

¹Accessed Feb. 26, 2022.

changes made to the data during the update: <https://www.rug.nl/ggdc/docs/pwt100-changelog.pdf>

Data file: “data/aggregates/pw91.dta” (.dta format)

Codebook: Available at <https://www.rug.nl/ggdc/productivity/pwt/pwt-releases/pwt9.1>

UN WPP

Data on country-level time series for child mortality were downloaded from United Nation World Population Prospects data (2019 revision) here: <https://population.un.org/wpp/>. The relevant file can be downloaded by selecting “Download Data Files”, then selecting the “CSV format” button and then choosing “Medium variant, annual, from 1950 to 2100 (CSV, 113.05 MB)” under the “Population by Age and Sex” subgroup. A copy of the data is provided in the replication directory.

Data file: “data/aggregates/WPP2019_Period_Indicators_Medium.csv” (.csv format)

Codebook: Available at <https://population.un.org/wpp/Download/Metadata/Documentation/>

Dataset list

Table 1: List of provided datasets

Data file	Source	Notes	Provided
data/aggregates/ucdp_prio/ucdp-prio-acd-201.dta	UCDP/PRIO	Public use	Yes
data/aggregates/pw91.dta	Penn World Table	Public use	Yes
data/aggregates/WPP2019_Period_Indicators_Medium.csv	UN WPP	Public use	Yes

Computational requirements

Software Requirements

INSTRUCTIONS: List all of the software requirements, up to and including any operating system requirements, for the entire set of code. It is suggested to distribute most dependencies together with the replication package if allowed, in particular if sourced from unversioned code repositories, Github repos, and personal webpages. In all cases, list the version *you* used.

- Stata (code was last run with version 17)
 - `estout` (as of 2021-02-17)
 - `tabout` (as of 2021-02-17)
 - the program “`executeallcode.do`” will install all dependencies locally.

- LaTeX Compiler

Controlled Randomness

- ☒ Random seed is set at line 1294 of program “03 analysis code.do” for the Monte Carlo simulations

Memory and Runtime Requirements

Summary Approximate time needed to reproduce the analyses on a standard (CURRENT YEAR) desktop machine:

- ☐ <10 minutes
- ☐ 10-60 minutes
- ☐ 1-8 hours
- ☐ 8-24 hours
- ☐ 1-3 days
- ☒ 3-14 days
- ☐ > 14 days
- ☐ Not feasible to run on a desktop machine, as described below.

Details The code was last run on a **4-core Intel-based laptop with MacOS version 10.15.7**.

Description of programs/code

All code is located in the “replicationdir/code” folder. After downloading the DHS data, the user must simply run the “replicationdir/code/excecuteallcode.do” STATA do file to replicate all data cleaning and analysis in the paper. In this section, we discuss the individual code files in more detail in case the user would like to modify analysis or data cleaning files.

The “excecuteallcode.do” file contains commands to automate the replication process by running 3 individual do files in the “replicationdir/code/dofiles” subfolder.

- “01 clean raw data.do”: cleans and appends raw DHS data
- “02 process intermediate data.do”: generates women-level and birth-level files and merges in non-DHS files
- “03 analysis code.do”: creates final analysis data and conducts all analysis from the paper

“01 clean raw data.do”

This STATA do file contains a data cleaning program, applies this program to each survey, and then appends the individual survey data into a single intermediate dataset. The code at the end of the file runs separate cleaning code do files for surveys that require data cleaning code due to differences in their data formats.

These separate cleaning files are stored in the relevant country data folders in “replicationdir/data/rawdhs”.

The main functions of the cleaning program are to:

- Clean data from sibling history module to create sibling mortality variables
- Clean data from birth history module to create child mortality variables
- Clean asset data to create a wealth index if this is not in the individual recode file
- For surveys in DHS round 4 or later: merge in the household member ID of the respondent’s co-resident mother (if applicable)

The data cleaning program has four arguments:

- Argument 1: The ID of the individual recode (IR) file
- Argument 2: The ID of the wealth index file, if it exists. Not all surveys have a separate wealth index file. If the wealth index data is in the individual recode, this argument is coded as “NA” for not applicable. For a few surveys, the wealth index is not provided and needs to be calculated. For this case, the argument is “CALC”.
- Argument 3: The ID of the person recode (PR) file. This argument is only relevant if argument 2 is the ID of wealth index file or “CALC”. The person recode is needed to merge information from the household-level file into the individual recode data.
- Argument 4: This file indicates the DHS Phase in which survey was conducted. The data cleaning steps depend on the DHS Phase. For one survey, the data format is different from the standard format from official phase reported by the DHS website which is reported in Tables 1 and 2 of this document. This survey is given an argument of “0”.

One important note: the code may produce an error during the data cleaning process as the DHS file names change if the DHS program has issued new versions of the data. These new versions are not frequent and usually the corrections to the data are minor. If this is the case, the user will need to manually correct the file names for the corresponding survey because the files names are used as arguments in the data cleaning program.

“02 process intermediate data.do” This STATA do file creates women and birth-level datasets from the intermediate data output written by “01 clean raw data.do”. It also merges in aggregate data from other sources (e.g. UN mortality data, GDP, conflict events) in “replicationdir/data/aggregates” to the birth-level dataset for the country x cohort panel analyses (Section 4.4. of the paper).

“03 analysis code.do” This STATA do file creates the final data files for analysis based on our sample restrictions and runs all analysis from the paper.

All tables and figures from this file are written into the “replicationdir/output” folder.

Note: Two of the appendix results are computationally intensive. Figure A.5 (“Robustness to Survey-by-Age Group Effects”) and A.6 (“Monte Carlo Simulations of Measurement Error”) may take several days to run on a personal laptop.

Instructions to Replicators

In order to replicate our analysis, the user will need to

- Step 1: Obtain permission access to data from the Demographic and Health Surveys (DHS) website. Other secondary datasets are provided in the replication directory.
- Step 2: Download DHS data
- Step 3: Run all cleaning and analysis code using “code/executeallcode.do”
- Step 4: Compile output into a single PDF file using the provided LaTeX document “compileoutput.tex”. This file will compile all tables and figures from the main text and Appendix of the paper.

Details

Step 1: Requesting access to DHS data

The user must register as a data user on the DHS website and download the raw data. Instructions to access DHS data are here: <https://dhsprogram.com/data/Access-Instructions.cfm>.²

The user should request access to data from all 44 countries in our analysis sample. See Tables 2 and 3 below for a complete list of surveys by country.

Step 2: Downloading DHS data

For each DHS survey, there are multiple data files types that contain information from each survey questionnaire. The files names for each questionnaire will have specific IDs (indicated in parenthesis below). Some file types are not available for all DHS surveys.

- Individual Recode (IR): main data file of interest; contains information from the women’s questionnaire
- Wealth Index (WI): contains household-level asset data information
- Person Recode (PR): household member roster data
- Household Recode (HH): household-level data
- Men’s Recode (MR): contains information from the men’s questionnaire

²Accessed Feb. 26, 2022.

More information on the DHS file formats and naming conventions is available here: <https://dhsprogram.com/data/File-Types-and-Names.cfm>

When downloading each data file type, the DHS will provide a folder with multiple files. You will only need the “DT” files that are in the STATA .DTA format. You will save these “DT” files directly to their corresponding country folder in the “replicationdir/data/rawdhs” directory.

Tables 2 and 3 provide the list of data files for each of the 119 DHS surveys in our analysis sample to download.

Step 3: Run “code/executeallcode.do”

Edit the directory path in “code/executeallcode.do” to match the user’s system. Run the provided “code/executeallcode.do” master STATA do file to run cleaning and analysis files.

Step 4: Compile output using “compileoutput.tex”

All output from the analysis code will be written into the “replicationdir/output” folder. The user can use the provided “compileoutput.tex” LaTeX document in the same folder to view all output in a single PDF document.

List of tables and programs

The provided code reproduces:

- ☐ All numbers provided in text in the paper
- ☒ All tables and figures in the paper
- ☐ Selected tables and figures in the paper, as explained and justified below.

Figure/Table	Program	Output file
Table 1	code/executeallcode.do	output/sumstat.tex
Figure 1	code/executeallcode.do	output/summary_plot_shares.pdf
Table 2	code/executeallcode.do	output/u5_anysibs_full_nostar; output/u5_numsibs_full_nostar
Table 3	code/executeallcode.do	output/covariates_full.tex; output/covariates_nonmigrants.tex; output/covariates_height.tex
Table 4	code/executeallcode.do	output/aggregates_nostar.tex
Table A.1	N/A	output/surveylist_countryyears_sample.tex
Table A.2	code/executeallcode.do	N/A
Table A.3	code/executeallcode.do	output/momdaughter_any.tex
Table A.4	code/executeallcode.do	output/momdaughter_count.tex
Table A.5	code/executeallcode.do	output/u5_bygender
Figure A.1	code/executeallcode.do	output/sibsize_mort.pdf
Figure A.2	code/executeallcode.do	output/summary_plot_logodds.pdf
Figure A.3	code/executeallcode.do	output/logit_by_age.pdf
Figure A.4	code/executeallcode.do	output/gradients.pdf
Figure A.5	code/executeallcode.do	output/surveyXagegroup.pdf
Figure A.6	code/executeallcode.do	data/simulation.dta; output/simulation.pdf
Figure A.7	code/executeallcode.do	output/dot_whiskers.pdf
Figure A.8	code/executeallcode.do	output/me_vs_or.pdf
Figure A.9	code/executeallcode.do	output/q5_series.pdf
Figure A.10	code/executeallcode.do	output/semiparametric_cells.pdf
Figure A.11	code/executeallcode.do	output/leave_one_out.pdf

Note: The statistics for Table A.2 are calculated in the script. The user can manually input them into the table in “compileoutput.do”

List of DHS surveys

Table 2: List of DHS data files to download

Country + Year	Type	Phase	Individual Recode (IR)	Wealth Index (WI)	Person Recode (PR)	Household Recode (HH)	Men's Recode (MR)
Afghanistan 2010	Special	VI	Y (see table notes)	N	N	Y (see table notes)	N
Afghanistan 2015	Standard	VII	Y	N	N	N	N
Bangladesh 2001	Special	IV	Y (see table notes)	N	N	Y	N
Benin 1996	Standard	III	Y	Y	Y	N	N
Benin 2008	Standard	V	Y	N	N	N	N
Bolivia 1994	Standard	III	Y	Y	Y	N	N
Bolivia 2003	Standard	IV	Y	N	N	N	N
Bolivia 2008	Standard	V	Y	N	N	N	N
Burkina Faso 1998-9	Standard	III	Y	Y	Y	N	N
Burkina Faso 2010	Standard	VI	Y	N	N	N	N
Burundi 2010	Standard	VI	Y	N	N	N	N
Burundi 2016-7	Standard	VII	Y	N	N	N	N
Cambodia 2000	Standard	IV	Y	Y	Y	N	N
Cambodia 2005	Standard	V	Y	N	N	N	N
Cambodia 2010	Standard	VI	Y	N	N	N	N
Cambodia 2014	Standard	VII	Y	N	N	N	N
Cameroon 1998	Standard	III	Y	Y	Y	N	N
Cameroon 2004	Standard	IV	Y	N	N	N	N
Cameroon 2011	Standard	VI	Y	N	N	N	N
Central African Republic 1994-5	Standard	III	Y	Y	Y	N	N
Chad 1996-7	Standard	III	Y	Y	Y	N	N
Chad 2004	Standard	IV	Y	N	N	N	N
Chad 2014-5	Standard	VII	Y	N	N	N	N
Congo 2005	Standard	V	Y	N	N	N	N
Congo 2011-2	Standard	VI	Y	N	N	N	N
Congo, Democratic Republic 2007	Standard	V	Y	N	N	N	N
Congo, Democratic Republic 2013-4	Standard	VI	Y	N	N	N	N
Côte d'Ivoire 1994	Standard	III	Y	Y	Y	N	N
Côte d'Ivoire 2011-2	Standard	VI	Y	N	N	N	N
Dominican Republic 2002	Standard	IV	Y	Y	Y	N	N
Dominican Republic 2007	Standard	V	Y	N	N	N	N
Ethiopia 2000	Standard	IV	Y	Y	Y	N	N
Ethiopia 2005	Standard	V	Y	N	N	N	N
Ethiopia 2011	Standard	VI	Y	N	N	N	N
Ethiopia 2016	Standard	VII	Y	N	N	N	N
Gabon 2000	Standard	IV	Y	Y	Y	N	N
Gabon 2012	Standard	VI	Y	N	N	N	N
Ghana 2007	Special	V	Y (see table notes)	N	N	Y (see table notes)	N
Guinea 1999	Standard	IV	Y	Y	Y	N	N
Guinea 2005	Standard	V	Y	N	N	N	N
Guinea 2012	Standard	VI	Y	N	N	N	N
Haiti 2000	Standard	IV	Y	Y	Y	N	N
Haiti 2005-6	Standard	V	Y	N	N	N	N
Haiti 2016-7	Standard	VII	Y	N	N	N	N
Indonesia 1994	Standard	III	Y	N	Y	N	N
Indonesia 1997	Standard	III	Y	Y	Y	N	N
Indonesia 2002-3	Standard	IV	Y	N	N	N	N
Indonesia 2007	Standard	V	Y	N	N	N	N
Indonesia 2012	Standard	VI	Y	N	N	N	N
Jordan 1997	Standard	III	Y	Y	Y	N	N
Kenya 1998	Standard	III	Y	Y	Y	N	N
Kenya 2003	Standard	IV	Y	N	N	N	N
Kenya 2008-9	Standard	V	Y	N	N	N	N
Kenya 2014	Standard	VII	Y	N	N	N	N
Lesotho 2004	Standard	IV	Y	N	N	N	N
Lesotho 2009	Standard	VI	Y	N	N	N	N
Lesotho 2014	Standard	VII	Y	N	N	N	N
Madagascar 1992	Standard	II	Y	N	Y	N	N
Madagascar 1997	Standard	III	Y	Y	Y	N	N
Madagascar 2003-4	Standard	IV	Y	N	N	N	N
Madagascar 2008-9	Standard	V	Y	N	N	N	N
Malawi 1992	Standard	II	Y	N	Y	N	Y (see table notes)
Malawi 2000	Standard	IV	Y	Y	Y	N	N
Malawi 2004	Standard	IV	Y	N	N	N	N
Malawi 2010	Standard	VI	Y	N	N	N	N
Malawi 2015-6	Standard	VII	Y	N	N	N	N

Notes: The IR and HH files are called “WOMEN.DTA” and “HOUSEHOLDS.DTA” for the Afghanistan 2010 data. The IR file has an of “IQ” in the Bangladesh 2001 and Ghana 2007 data. The household survey data for Ghana 2007 is split into two files “GHHA51FL.DTA” and “GHHB51FL.DTA”. Both are used in data cleaning. The Malawi 1992 data is the only survey where the Men’s Recode is used for data cleaning.

Table 3: List of DHS data files to download (continued)

Country + Year	Type	Phase	Individual Recode (IR)	Wealth Index (WI)	Person Recode (PR)	Household Recode (HH)	Men's Recode (MR)
Mali 1995-6	Standard	III	Y	Y	Y	N	N
Mali 2001	Standard	IV	Y	Y	Y	N	N
Mali 2006	Standard	V	Y	N	N	N	N
Mali 2012-3	Standard	VI	Y	N	N	N	N
Morocco 1992	Standard	II	Y	Y	Y	N	N
Morocco 2003-4	Standard	IV	Y	N	N	N	N
Mozambique 1997	Standard	III	Y	Y	Y	N	N
Mozambique 2003	Standard	IV	Y	N	N	N	N
Mozambique 2011	Standard	VI	Y	N	N	N	N
Namibia 1992	Standard	II	Y	Y	Y	N	N
Namibia 2000	Standard	IV	Y	Y	Y	N	N
Namibia 2013	Standard	VI	Y	N	N	N	N
Nepal 1996	Standard	III	Y	Y	Y	N	N
Nepal 2006	Standard	V	Y	N	N	N	N
Nepal 2016	Standard	VII	Y	N	N	N	N
Niger 1992	Standard	II	Y	N	Y	N	N
Niger 2006	Standard	V	Y	N	N	N	N
Niger 2012	Standard	VI	Y	N	N	N	N
Nigeria 2008	Standard	V	Y	N	N	N	N
Nigeria 2013	Standard	VI	Y	N	N	N	N
Peru 1991-2	Standard	II	Y	Y	Y	N	N
Peru 1996	Standard	III	Y	Y	Y	N	N
Peru 2000	Standard	IV	Y	Y	Y	N	N
Philippines 1993	Standard	III	Y	Y	Y	N	N
Philippines 1998	Standard	III	Y	Y	Y	N	N
Rwanda 2000	Standard	IV	Y	Y	Y	N	N
Rwanda 2005	Standard	V	Y	N	N	N	N
Rwanda 2010	Standard	VI	Y	N	N	N	N
Rwanda 2014-5	Standard	VII	Y	N	N	N	N
São Tomé & Príncipe 2008-9	Standard	V	Y	N	N	N	N
Senegal 1992-3	Standard	II	Y	N	Y	N	N
Senegal 2005	Standard	IV	Y	N	N	N	N
Senegal 2010-1	Standard	VI	Y	N	N	N	N
Sierra Leone 2008	Standard	V	Y	N	N	N	N
Sierra Leone 2013	Standard	VI	Y	N	N	N	N
South Africa 1998	Standard	III	Y	Y	Y	N	N
South Africa 2016	Standard	VII	Y	N	N	N	N
Swaziland (Eswatini) 2006-7	Standard	V	Y	N	N	N	N
Tanzania 1996	Standard	III	Y	Y	Y	N	N
Tanzania 2004-5	Standard	IV	Y	N	N	N	N
Tanzania 2010	Standard	VI	Y	N	N	N	N
Tanzania 2015-6	Standard	VII	Y	N	N	N	N
Togo 1998	Standard	III	Y	Y	Y	N	N
Togo 2013-4	Standard	VI	Y	N	N	N	N
Zambia 1996	Standard	III	Y	Y	Y	N	N
Zambia 2001-2	Standard	IV	Y	Y	Y	N	N
Zambia 2007	Standard	V	Y	N	N	N	N
Zambia 2013-4	Standard	VI	Y	N	N	N	N
Zimbabwe 1994	Standard	III	Y	Y	Y	N	N
Zimbabwe 1999	Standard	IV	Y	Y	Y	N	N
Zimbabwe 2005-6	Standard	V	Y	N	N	N	N
Zimbabwe 2010-1	Standard	VI	Y	N	N	N	N
Zimbabwe 2015	Standard	VII	Y	N	N	N	N

References

- Feenstra, Robert C, Robert Inklaar, and Marcel P Timmer (2015) “The Next Generation of the Penn World Table,” *American Economic Review*, 105(10), 3150-3182, available for download at www.ggdc.net/pwt.
- Gleditsch, Nils Petter, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg, and Håvard Strand (2002) “Armed conflict 1946-2001: A new dataset,” *Journal of peace research*, 39 (5), 615–637.
- ICF (2004-2017) “Demographic and Health Surveys (various) [Datasets].,” *Funded by USAID. Rockville, Maryland: ICF [Distributor]*.
- Pettersson, Therese, Shawn Davies, Amber Deniz, Garoun Engström, Nanar Hawach, Stina Höglbladh, and Margareta Sollenberg Magnus Öberg (2021) “Organized violence 1989–2020, with a special emphasis on Syria,” *Journal of Peace Research*, 58 (4), 809–825.
- United Nations, Department of Economic and Population Division Social Affairs (2019) “World Population Prospects 2019: Data Booklet (ST/ESA/SER.A/424).”
-