README

Replication Package

"Consequences of Growing Up Displaced: Evidence from Displaced Children in Colombia"
VenRePs-Kids 2025

Overview

This repository contains all files necessary to replicate the analysis in *"Consequences of Growing Up Displaced: Evidence from Displaced Children in Colombia."* The package is organized into two main components: **build** (dataset construction) and **analysis** (tables and figures). The analysis uses Stata as the primary software, with R for one appendix figure and Python (Jupyter) for another.

Approximate run time:

- Stata: approximately 3.5 minutes
- R: less than 1 minute
- Python: approximately 1 minute

Data Availability Statement

The analysis relies on primary data from the *Venezuelan Repatriates Kids* (VenRePs-Kids) survey, conducted in two waves in Colombia (Wave 1 and Wave 2). These data were collected by the authors and are included in this replication package under `build/input/`. No external or restricted data sources are required to reproduce the results.

All input data files are included in the package. No registration, application, or third-party approval is required to access them.

Software Requirements

Stata

- Version: StataNow/SE 19.5
- Required user-written packages:
  - estout (provides esttab, eststo) — table export
  - coefplot — coefficient plots
  - omegacoef — calculation of McDonald's omega coefficient (internal consistency reliability of scales)

- mdesc — summary and description of missing values in variables

Install packages by running:

```
ssc install estout
ssc install coefplot
ssc install omegacoef
ssc install mdesc
```

R

- Version: 4.4.2
- Required packages: tidyverse, haven, sf, viridis, ggspatial, ggplot2, grid

Install packages by running:

```
install.packages(c("tidyverse", "haven", "sf", "viridis", "ggspatial", "ggplot2",
"grid"))
```

Python

- Version: 3.13 (tested on 3.13.11)
- Environment: Anaconda
- Development environment: Jupyter Notebook executed in Visual Studio Code
- Required packages: pandas, geopandas, pyreadstat, osmnx, matplotlib

Install packages by running:

```
!pip install pandas geopandas pyreadstat osmnx matplotlib
```

Instructions for Replicators

Step 1 — Set the working directory

Open `analysis/code/master.do` and update the global `wd` to the path where you saved this replication package on your computer:

```
global wd "/your/local/path/Final RP"
```

This is the **only path you need to change.** All other paths are set automatically relative to `wd`.

Step 2 — Run the master Stata script

Execute `analysis/code/master.do` in Stata. This script:

- Sets all global directory paths.
- Runs
- Runs `build/code/data_creation.do` and `build/code/imputation_process.do` to construct the analysis datasets.
- Runs all paper and appendix Stata analysis scripts in order.

Step 3 — Run the R script (Appendix A map)

Open `analysis/code/appendix/2.1_appendix_A.R` in R and update `project_dir` to match your local path:

```
project_dir <- "/your/local/path/Final RP"
```

Then run the full script. Output: `Figure_A1_a.png` saved to `analysis/outputs/appendix/`.

Step 4 — Run the Jupyter notebook (Appendix B map)

Open `analysis/code/appendix/2.2_appendix_B.ipynb` in Jupyter and update `PROJECT_DIR` in the workspace cell:

```
PROJECT_DIR = Path("/your/local/path/Final RP")
```

Then run all cells. Outputs: `Figure_B2.png`, `Figure_B2.pdf`, `Figure_B2_status.png`, `Figure_B2_status.pdf` saved to `analysis/outputs/appendix/`.

> **Note:** The notebook downloads geographic boundary data from OpenStreetMap on the first run (requires internet connection).

Dataset Descriptions

Primary Survey Data — VenRePs-Kids (build/input/)

File	Description
VenReps_Kids_Variables_Final.dta	Child-level variables, Wave 1
VenReps_Kids_Variables_Final_W2.dta	Child-level variables, Wave 2
VenReps_HH_Variables.dta	Household-level variables — used for wealth index construction
VenReps_HH_Variables_Final.dta	Household-level variables — cleaned, used in analysis
VenReps_MM_Variables.dta	Mother/caregiver variables — raw
VenReps_MM_Variables_Final.dta	Mother/caregiver variables — cleaned, used in analysis
wealth_i_wave1.dta	Wealth index inputs, Wave 1
wealth_i_wave2.dta	Wealth index inputs, Wave 2
pca_params_w1.dta	PCA loadings, means, and standard deviations from Wave 1 — applied to Wave 2 for comparability
Coordinates_VenReps_kids_final_baseline.dta	GPS coordinates of surveyed households (baseline)

Supplementary Datasets (build/input/)

File	Description
bmi_zscore_all.xlsx	WHO reference tables for BMI z-score calculation

File	Description
victimization_dta.dta	Victimization history data merged into the main dataset
ruv_merge.dta	Registro Único de Víctimas (RUV) — official Colombian forced displacement registry
flujos_migracion_colombia.xlsx	Municipal-level Venezuelan migration flows into Colombia

Geographic Files (build/input/MGN2018_MPIO_POLITICO/)

File	Description
MGN_MPIO_POLITICO.dta (and .shp, .shx, .dbf, .prj)	DANE 2018 Colombian municipal political boundaries — shapefile used for maps
ven_muni_count.dta	Venezuelan population counts by municipality, 2018 Colombian Census

Build: Dataset Construction (build/code/)

These scripts are called automatically by `master.do`. They must be run before any analysis.

Script	Purpose	Main Outputs
data_creation.do	Constructs the child-level analytical panel. Harmonizes Wave 1 and Wave 2 data; builds pre- and post-migration wealth indices via PCA; merges child, household, caregiver, and victimization records; creates displacement status indicators (non-displaced Colombian, IDP, Venezuelan refugee); standardizes developmental, mental health, and anthropometric outcomes by age group; constructs rank-based variables for mobility and gradient analyses.	wave1.dta, wave2.dta, panel_2026.dta
imputation_process.do	Constructs the attrition-adjusted panel under three scenarios: observed data, mean imputation, and Lee bounds (lower and upper). Stacks all scenarios into a single long-format analytical dataset.	panel_imputed.dta

Constructed Datasets (build/output/)

File	Description
wave1.dta	Main analysis dataset, Wave 1
wave2.dta	Main analysis dataset, Wave 2

File	Description
panel_2026.dta	Pooled panel (Waves 1 and 2), used for DiD and gradient analyses
panel_imputed.dta	Panel with imputed observations for attrition robustness checks (Appendix F)
wealth_i_wave1.dta	Constructed wealth index, Wave 1
Wealth_i wave2.dta	Constructed wealth index, Wave 2

Analysis: Tables and Figures

Paper Content (analysis/code/paper/)

Script	Outputs
1.1_paper_descriptives.do	Table 1, Table 2, Figure 1, Figure 2
1.2_paper_gaps_and_mobility.do	Table 3, Figure 3 (all outcomes), Figure 4

Appendix Content (analysis/code/appendix/)

Script	Language	Appendix	Outputs
2.1_appendix_A.do	Stata	Appendix A	Figure A1(b), Figure A2
2.1_appendix_A.R	R	Appendix A	Figure A1(a) — choropleth map, Venezuelan distribution 2018
2.2_appendix_B.do	Stata	Appendix B	Tables B2–B11, Figure B2 (status)
2.2_appendix_B.ipynb	Python	Appendix B	Figure B2 — household location map, Valle de Aburrá
2.3_appendix_C.do	Stata	Appendix C	Tables C1–C7, Figures C1–C4
2.4_appendix_D.do	Stata	Appendix D	Table D25, Tables D (heterogeneity by subgroup), Figures D1–D3
2.5_appendix_E.do	Stata	Appendix E	Tables E1–E2
2.6_appendix_F.do	Stata	Appendix F	Tables F (pooled and DiD specifications, all outcomes)

List of Exhibits

Main Paper

Exhibit	Script	Output File
Table 1	1.1_paper_descriptives.do	outputs/paper/Table1.tex
Table 2	1.1_paper_descriptives.do	outputs/paper/Table2.tex
Table 3	1.2_paper_gaps_and_mobility.do	outputs/paper/Table3.tex
Figure 1	1.1_paper_descriptives.do	outputs/paper/Figure1.png
Figure 2	1.1_paper_descriptives.do	outputs/paper/Figure2.png
Figure 3 (BMI)	1.2_paper_gaps_and_mobility.do	outputs/paper/Figure3_bmi.png
Figure 3 (Depression)	1.2_paper_gaps_and_mobility.do	outputs/paper/Figure3_dep.png
Figure 3 (Anxiety)	1.2_paper_gaps_and_mobility.do	outputs/paper/Figure3_anx.png
Figure 3 (SDQ)	1.2_paper_gaps_and_mobility.do	outputs/paper/Figure3_sdq.png
Figure 3 (Problem behaviors)	1.2_paper_gaps_and_mobility.do	outputs/paper/Figure3_pb.png
Figure 3 (Trauma — TSCYC)	1.2_paper_gaps_and_mobility.do	outputs/paper/Figure3_tscyc.png
Figure 4	1.2_paper_gaps_and_mobility.do	outputs/paper/Figure4.png

Appendix

Exhibit	Script	Output File
Figure A1(a) — Migrant Distribution Map	2.1_appendix_A.R	outputs/appendix/Figure_A1_a.png
Figure A1(b)	2.1_appendix_A.do	outputs/appendix/Figure_A1_b.png
Figure A2	2.1_appendix_A.do	outputs/appendix/Figure_A2.png
Figure B2 (by nationality)	2.2_appendix_B.ipynb	outputs/appendix/Figure_B2.png
Figure B2 (by status)	2.2_appendix_B.ipynb	outputs/appendix/Figure_B2_status.png
Figures C1–C4	2.3_appendix_C.do	outputs/appendix/Figure_C1.png – Figure_C4.png
Figures D1–D3	2.4_appendix_D.do	outputs/appendix/Figure_D1.png – Figure_D3.png
Tables B2–B11	2.2_appendix_B.do	outputs/appendix/Table_B2.tex – Table_B11.tex
Tables C1–C7	2.3_appendix_C.do	outputs/appendix/Table_C1.tex – Table_C7.tex

Exhibit	Script	Output File
Tables D (heterogeneity)	2.4_appendix_D.do	outputs/appendix/Tables_D_[outcome]_[subgroup].tex
Table D25	2.4_appendic_D.do	outputs/appendix/Table_D25.tex
Tables E1–E2	2.5_appendix_E.do	outputs/appendix/Table_E1.tex, Table_E2.tex
Tables F (pooled & DiD)	2.6_appendix_F.do	outputs/appendix/Table_F_[outcome]_[pooled/did].tex

Outcomes covered in Tables D and F: `bmi, dep, anx, sdq, pb, trauma, mh.`

Subgroups in Tables D: `enroll, insurance, sisben, status.`

Folder Structure

```
Final RP/
├── analysis/
│   ├── code/
│   │   ├── master.do                    ← Run this first (Stata)
│   │   ├── paper/
│   │   │   ├── 1.1_paper_descriptives.do
│   │   │   └── 1.2_paper_gaps_and_mobility.do
│   │   └── appendix/
│   │       ├── 2.1_appendix_A.do
│   │       ├── 2.1_appendix_A.R         ← Run separately in R
│   │       ├── 2.2_appendix_B.do
│   │       ├── 2.2_appendix_B.ipynb     ← Run separately in Jupyter
│   │       ├── 2.3_appendix_C.do
│   │       ├── 2.4_appendix_D.do
│   │       ├── 2.5_appendix_E.do
│   │       └── 2.6_appendix_F.do
│   └── outputs/
│       ├── paper/                       ← Tables and figures for main text
│       └── appendix/                    ← Tables and figures for appendix
└── build/
    ├── code/
    │   ├── data_creation.do
    │   └── imputation_process.do
    ├── input/                           ← Raw and supplementary data (do not modify)
    └── output/                          ← Constructed analysis datasets
```

Contact

For questions about the replication package, please contact the authors at:
- sandrarozo@worldbank.org
- hiller@ucdavis.edu
- a.moya@uniandes.edu.co