

Reproducibility Package

Public Externalities of Universal Health Care Access

Tatiana Hiller (University of California, Davis) · Leonardo Peñaloza-Pacheco (Cornell University) · Sandra V. Rozo (World Bank)

1. Overview

This package reproduces every table and figure in the paper Public Externalities of Universal Health Care Access. The paper studies Colombia's 2022 Estatuto Temporal de Protección para Migrantes Venezolanos (ETPV), which granted roughly two million undocumented Venezuelan migrants full access to the health system, and estimates its effect on mortality among the Colombian host population using a municipality-level difference-in-differences design.

The code is written in Stata (analysis) with short Python scripts (three assemble table fragments; two prepare the birth and age-population intermediates). Running the single master script `main.do` reproduces all 16 tables and 33 figures. Two small aggregate intermediate files for the birth and age-specific-population denominators are provided in `code/temp/` as usable data (see Section 2.3), with the scripts that reproduce them from the raw DANE files included in `code/python/`.

Data availability summary: ☒ Some data cannot be made publicly available.

Several inputs (RIPS health-service records, EEVV vital statistics, ETPV registrations, PAI vaccination records) are confidential administrative data that cannot be made publicly available and must not be redistributed or posted.

2. Data Availability Statement

☒ Some data cannot be made publicly available.

☒ The authors certify that they had legitimate access to and permission to use all data used in the manuscript.

Confidential datasets were accessed through the authors' institutional access to Colombia's Sistema Integral de Información de la Protección Social (SISPRO) and through agreements with Migración Colombia; they cannot be made publicly available or redistributed.

Restricted data (not for public release)

Filename (in data/)	Content	Source	Access
<code>rips_data/rips_detailed_panel.csv</code>	Migrant health-service encounters by municipality × month × document type × care type	Ministry of Health, RIPS via SISPRO cube	Restricted. SISPRO cube access (institutional). https://www.sispro.gov.co . Accessed 2026
<code>rips_data/rips_colombian_panel.csv</code>	Colombian citizens' service utilization by municipality × month × care type	Ministry of Health, RIPS via SISPRO cube	Restricted. As above. Accessed 2026
<code>ETPV_MunicipalityDATA.dta</code>	ETPV registrations and PPT deliveries by municipality (the treatment measure)	Migración Colombia (administrative)	Restricted. By agreement with Migración Colombia. Accessed 2026
<code>PAI/pai_vacunacion_panel.dta</code>	Vaccine doses applied and programmatic targets by	Ministry of Health, PAIWEB via SISPRO	Restricted. SISPRO access. Accessed 2026

Filename (in data/)	Content	Source	Access
	municipality × year × vaccine		
rips_data/rips_disease_unified_panel.csv	RIPS diagnoses grouped into disease categories (used only for the chronic-care aggregate rips_dis_m.dta)	Ministry of Health, RIPS via SISPRO cube	Restricted. As above. Aggregated by agg_disease.py; the resulting non-confidential aggregate ships as usable data. Accessed 2026

The three restricted inputs can be requested through the routes below. These are the routes the authors used themselves in 2026, including the formal petition route under Colombian law, which is the reliable channel when an informal request goes unanswered.

RIPS (Registro Individual de Prestación de Servicios de Salud) and PAIWEB. Holder: Ministerio de Salud y Protección Social (MinSalud), delivered through the SISPRO analytical cubes. Route: institutional request to MinSalud for SISPRO cube access; the aggregated cubes are reachable at <https://www.sispro.gov.co> once credentials are issued, and the municipality-by-month extracts used here are produced from those cubes. Contact: correo@minsalud.gov.co (official correspondence address); acknowledgements are issued by notificaciones_sgdea@minsalud.gov.co. Formal route if the informal one stalls: a derecho de petición under Article 23 of the Colombian Constitution and Law 1755 of 2015, sent to the same address; MinSalud returns a radicado number and must answer within fifteen business days (the authors' petition of 19 August 2026 was assigned radicado 2026423003246692). A replicator outside Colombia may need a Colombian host institution. Individual-level records are not released; the municipal aggregates used here are.

ETPV registrations and PPT deliveries. Holder: Unidad Administrativa Especial Migración Colombia. Route: the PQRSDF portal is the only working channel — <https://apps.migracioncolombia.gov.co/pqrs/home>. Migración does not answer data requests by email. File the request as a derecho de petición de interés general; the portal issues a radicado immediately and the statutory window is fifteen business days (the authors' request of 22 August 2026 is radicado PQRSDF N° 332450). Ask specifically for registrations and permits by DIVIPOLA municipality code and month: a request phrased by municipality name alone returns a name-keyed file, which cannot be merged safely because 67 Colombian municipality names are shared across departments.

Note on the ETPV merge. Because the extract the authors received was keyed by municipality name, the package resolves names to DIVIPOLA codes in code/python/assign_etpv_codes.py and validates the result against Migración Colombia's public open-data snapshot (which carries name and department). The validation prints on every run: the correlation between the municipal permit counts and the snapshot is 0.999999 across the 924 municipalities whose names are unique in both sources.

Public data (freely downloadable from the sources below)

Filename (in data/)	Content	Source	URL / access
mortality/NO_FETALES/*.dta, nofetal2021.csv	DANE vital statistics (EEVV) non-fetal death microdata, 2016–2024	DANE (Estadísticas Vitales)	https://microdatos.dane.gov.co (Estadísticas Vitales – EEVV). Accessed 2026
evv_births/nacimientos_panel_municipio_mes.dta	DANE birth records, aggregated to municipality × month	DANE (EEVV, births)	https://microdatos.dane.gov.co . Accessed 2026
births/nacimientos_*.xls, nacimientos_*_c3.xls	DANE annual birth microdata	DANE (EEVV, births)	https://microdatos.dane.gov.co . Files for 2019–2023 are OLE-encrypted with Excel's default password

Filename (in data/)	Content	Source	URL / access
	(Cuadro 3), 2015–2024; used for the age-specific denominators		(VelvetSweatshop); decrypt_births.py opens them. Accessed 2026
population/DANE_pop_sexo_edad_Mun_2005_2019.xlsx, DANE_pop_sexo_edad_Mun_2020_2035.xlsx	Municipal population by age group	DANE	https://www.dane.gov.co (Proyecciones de población). Accessed 2026
population/DANE_proypoblacion_Mun_2020_2035.xlsx, DANE_retroproypoblacion_Mun_2005_2019.xlsx	Municipal population projections and retroprojections by age	DANE	https://www.dane.gov.co (Proyecciones de población). Accessed 2026
cede/cede_master_panel.dta	Pre-shock municipal characteristics	CEDE, Universidad de los Andes	https://datoscede.uniandes.edu.co . Accessed 2026
growth/Cleaning_PEP_GEIH_adm2_2010_2024.dta	Labor-market cells built from the household survey	DANE, GEIH microdata	https://microdatos.dane.gov.co (GEIH). Accessed 2026
growth/Cleaning_PEP_colombia_adm2.dta	Municipal night-light density and fiscal accounts	VIIRS (NOAA/Colorado School of Mines) and municipal fiscal accounts (DNP)	https://eogdata.mines.edu/products/vnl/ . Accessed 2026

The DANE microdata catalogue does not expose stable deep links to individual study pages, and study identifiers change between releases, so the catalogue plus the exact study name is given rather than links that will break:

- EEVV births microdata: <https://microdatos.dane.gov.co> , catalogue entry "Estadísticas Vitales - Nacimientos", annual files 2016–2024.
- EEVV deaths microdata (non-fetal): <https://microdatos.dane.gov.co> , "Estadísticas Vitales - Defunciones No Fetales", 2016–2024.
- GEIH: <https://microdatos.dane.gov.co> , "Gran Encuesta Integrada de Hogares", annual files 2010–2024.
- Population projections and retroprojections: <https://www.dane.gov.co> , Demografía y Población → Proyecciones de Población, municipal series 2005–2019 (retroproyecciones) and 2020–2035.
- CEDE Panel Municipal: <https://datoscede.uniandes.edu.co> , Panel Municipal (free registration required).
- VIIRS night lights: <https://eogdata.mines.edu/products/vnl/> , annual composites.

Additional public input used for validation. Migración Colombia's public open-data snapshot (municipality name and department) is read by code/python/assign_etpv_codes.py to validate the DIVIPOLA code assignment of the ETPV extract (see the merge note in Section 2.1). It is publicly available.

2.1 Raw source files

The DANE annual birth cuadros (births/nacimientos_*.xls, nacimientos_*_c3.xls, 2015–2024) and the two municipal population-by-age files (population/DANE_pop_sexo_edad_Mun_2005_2019.xlsx, DANE_pop_sexo_edad_Mun_2020_2035.xlsx) are included in the package. With the cuadros in data/births/ and the population files in data/population/, the two optional preparation scripts (decrypt_births.py, then parse_births_popage.py) run end-to-end and regenerate the aggregates in code/temp/ that the main analysis reads.

Caution for the metadata: DANE reissues these cuadros as vital statistics are consolidated, so a copy downloaded later may differ in the most recent year or two from the copy behind the published results. The files shipped here are the ones used for the paper; if the archive keeps a fixed copy, these are the ones to keep.

2.2 Derived and combined public files (construction notes)

Four of the public files above are not raw source files: one is derived and three are combinations of several sources. The full per-file detail below is reproduced from the authors' supplement `derived_data_notes/README.md`. One entry per file, giving its status, the underlying sources, and how it is constructed.

evv_births/nacimientos_panel_municipio_mes.dta

Status: derived.

Underlying source. DANE, Estadísticas Vitales (EEVV), nacidos vivos. Birth-certificate microdata, annual files 2016 to 2024. Catalogue: <https://microdatos.dane.gov.co>, entry "Estadísticas Vitales - Nacimientos".

Construction. Individual birth records are collapsed to municipality of residence (`cod_dpto_res`, `cod_mun_res`) by month (`anno`, `mes`).

- Continuous fields become a mean and a count of non-missing records: prenatal visits (`consultas_mean`, `consultas_n`), Apgar at one and five minutes, father's age, live births to the mother, number of pregnancies.
- Every categorical field becomes a set of counts, one variable per category: birth weight band (`n_peso_*`), length band (`n_talla_*`), gestational age band in both the raw and the CIE grouping (`n_tges_*`, `n_tges_cie_*`), mother's age band (`n_edadm_*`), sex, delivery type, multiplicity, place of delivery (`n_sitparto_*`), attendant (`n_atenpar_*`), insurance regime (`n_segso_*`), mother's education (`n_nivedum_*`), area of birth and of residence, ethnicity, marital status, blood group and Rh, and mother's country of residence (`n_codpres_*`, `n_madre_venezolana`).
- Derived indicators are cut from those bands: `n_low_bw` and `n_very_low_bw` (under 2,500g and under 1,500g), `n_macrosomia`, `n_preterm` and `n_very_preterm`, `n_post_term`, the prenatal-visit groupings (`n_prenatal_0`, `n_prenatal_1_3`, `n_prenatal_4_6`, `n_prenatal_7plus`), `n_apgar1_low` and `n_apgar5_low`, and the maternal-age flags.

Caution for a replicator. The birth certificate records weight and gestation in bands, not in grams and weeks. Low birth weight is therefore the union of the bands below 2,500g, and preterm is the CIE grouping's bands below 37 weeks. Sub-32-week preterm is not recoverable from the banding; use under 28 weeks if a finer cut is needed. The "no information" bands are excluded from denominators rather than treated as zero.

Coverage note. Country of birth of the mother (`CODPAISNACMAD`) exists only from 2024. The composition statistics in Section 5 and Appendix F that use it are computed directly from the 2024 microdata by `code/65b_composition_2024.do`, not from this panel.

cede/cede_master_panel.dta

Status: combined.

Underlying source. CEDE Panel Municipal, Centro de Estudios sobre Desarrollo Económico, Universidad de los Andes. <https://datoscede.uniandes.edu.co> (free registration required). CEDE distributes the panel as separate thematic modules, among them *características generales*, *buen gobierno*, *educación*, *salud*, *agrícola*, *conflicto y violencia*, and the fiscal series.

Construction. The modules are merged on municipality code (`codmpio`) and year (`ano`) into one wide panel. Two bookkeeping variables survive from that merge and make it auditable: `n_panels_present` records how many source modules contributed to each municipality-year, and `n_vars_nonmissing` how many variables are populated.

What the paper uses. Only thirteen pre-shock variables, all measured before 2020 and listed in the note to Table 3: poverty, health-access and insurance deprivation, rurality, altitude, distance to the department capital and to Bogotá, water and sewerage coverage, state capacity, health transfers per capita, homicides, and displaced-population inflows. The panel ends in 2021, which is why it serves as a pre-shock characteristic set rather than a control that could respond to the program.

growth/Cleaning_PEP_GEIH_adm2_2010_2024.dta

Status: combined.

Underlying sources. DANE, Gran Encuesta Integrada de Hogares (GEIH), person-level microdata pooled 2010 to 2024 (<https://microdatos.dane.gov.co>, entry "Gran Encuesta Integrada de Hogares"), merged with municipal covariates drawn from CEDE and from VIIRS night lights.

Construction. GEIH person records retaining the survey identifiers (`directorio`, `secuencia_p`, `orden`, `hogar`), the demographic and labour variables (employment, hours, formality, contract type, earnings, industry and occupation codes, country of birth), and the survey weight, with the municipality code (`adm2_pcode`) attached. Pre-period municipal covariates are merged on: 2009 GDP per capita, 2005 poverty and NBI, 2010 theft and homicide rates, baseline night lights, and the 2010 urban share.

Provenance note. The treatment variables carried in the file (`dtreatment`, `dpost`, `post_cont_*`, and the percentile variants `dtreatp5` through `dtreatp75`) were constructed for the earlier PEP regularization project from which this file comes. **The present paper does not use them.** It uses the labour variables and the municipality identifier only.

growth/Cleaning_PEP_colombia_adm2.dta

Status: combined.

Underlying sources.

- VIIRS night lights, annual composites, Earth Observation Group, Colorado School of Mines. <https://eogdata.mines.edu/products/vnl/>
- Municipal fiscal accounts (ejecuciones presupuestales), Departamento Nacional de Planeación.
- Municipal characteristics from CEDE, and emissions and inequality series.

Construction. A municipality-by-year panel merging annual mean night-light density (`light_mean`, `n1_meanyear*`) with municipal revenue and expenditure aggregates (the `y_*` and `g_*` families, in levels, per capita and in logs) and baseline characteristics (poverty, NBI, Gini, urban share, population).

Provenance note. As with the GEIH file, this was assembled for the earlier PEP project, and it carries that project's treatment indicators (`tpep`, `dtreatment`, `dtreat_post`). **The present paper uses the night-light and municipal fiscal series only**, in the local-activity test reported in the mechanisms section.

The two `Cleaning_PEP_*` files were assembled for the earlier PEP project; that project's assembly code is not part of this package, and the files are provided here as documented inputs, with sources and construction as above.

Where these files are read

File	Read by
<code>nacimientos_panel_municipio_mes.dta</code>	<code>code/45_births.do</code> (Appendix F)
<code>cede_master_panel.dta</code>	<code>code/44_cede_baseline.do</code> , then the robustness battery
<code>Cleaning_PEP_GEIH_adm2_2010_2024.dta</code>	<code>code/47_labor_geih.do</code> , <code>code/49_health_workforce.do</code> , <code>code/50_mech_evidence.do</code> , <code>code/51_pyramid.do</code> , <code>code/65_data_appendix.do</code>
<code>Cleaning_PEP_colombia_adm2.dta</code>	<code>code/46_nightlights.do</code>

2.3 Provided intermediate data (usable data)

Two small, fully aggregate intermediate files are shipped in `code/temp/` so the package runs without the fragile step of parsing dozens of encrypted DANE birth spreadsheets:

File	Content	Produced by
<code>code/temp/births_muni_year.csv</code>	Live births by municipality × year (infant/maternal denominators)	<code>code/python/parse_births_popage.py</code>
<code>code/temp/pop_age_muni_year.csv</code>	Population by municipality × year × age band	<code>code/python/parse_births_popage.py</code>
<code>code/temp/rips_dis_m.dta</code>	Colombian chronic-care consultation counts by municipality × month (for the mediation figure)	<code>code/python/agg_disease.py</code> (from the RIPS disease panel)

Both are non-confidential aggregates (municipality-year counts). To regenerate them from the raw DANE files instead of using the provided versions, run `python code/python/decrypt_births.py` then `python code/python/parse_births_popage.py` (see Requirements for the additional Python packages this needs). The main run uses the provided files.

Note on the `code/temp/` folder (added from the authors' response, August 2026). **Stata and Python write intermediate files into `code/temp/` during the run.** The package ships only three pre-computed usable-data files there — the ones that require restricted inputs to regenerate (documented above in this section):

- `births_muni_year.csv` — municipal birth counts by year (from the encrypted DANE births).
- `pop_age_muni_year.csv` — municipal population by age group and year.
- `rips_dis_m.dta` — municipal disease-category counts (from the confidential RIPS).

Everything else in the folder is regenerated by `main.do`.

3. Instructions for Replicators

1. All input datasets listed in Section 2 are already in place under `data/` in this internal-verification copy (`data/rips_data/`, `data/mortality/NO_FETALES/`, `data/population/`, `data/cede/`, `data/growth/`, `data/evv_births/`, `data/PAI/`, and `ETPV_MunicipalityDATA.dta` in `data/`). Nothing needs to be added.
2. Open `main.do`. The project global at the top is already set to this folder; if you move the folder, update that one line to its new full path:

```
global project "/your/path/to/Reproducibility Package/package"
```
3. Run `main.do` in full. It installs the required Stata commands into `code/ado`, runs the analysis from raw data, and writes every exhibit to `tables/` and `figures/`. The three Python steps at the end (standard library only) assemble the mechanism and robustness tables.
4. No other file or path needs editing.

4. List of Exhibits

Every table and figure in the paper, its output file, and the script that generates it. Table files are `.tex` fragments in `tables/`; figures are `.png` in `figures/`. Tables marked `make_*.py` are assembled by the Python step from Stata-produced fragments. Three appendix tables (`tab_causes_wrap`, `tab_causecap_wrap`, `tab_births_wrap`) are authored caption/notes shells shipped as source; the code regenerates the numeric fragments they include.

Exhibit	Output file	Generating script
Figure 1	<code>figures/fig_a5_permit_ratio.png</code>	<code>code/16_fig_permit_ratio.do</code>
Figure 2	<code>figures/fig_a3_combined_shares.png</code>	<code>code/7_fig_a3_combined_shares.do</code>

Exhibit	Output file	Generating script
Figure 3	figures/fig_migrant_caretype.png	code/69_fig_migrant_caretype.do
Figure 4	figures/fig_muni_first_stage.png	code/5_fig_a4_muni_first_stage.do
Figure 5	figures/fig_es_nationality.png	code/37_event_studies_nationality.do
Figure 6	figures/fig_a19_event_studies.png	code/27_event_studies.do
Table 1	tables/tab_did_mortality.tex	code/31_tables_did.do
Table 2	tables/tab_robustness2.tex	code/python/make_robustness2_table.py
Table 3	tables/tab_robustness.tex	code/python/make_robustness_table.py
Table 4	tables/tab_mech_growth_wrap.tex	code/python/make_mech_tables.py
Table 5	tables/tab_decomp.tex	code/61_table_decomp.do
Table 6	tables/tab_care_switch.tex	code/62_care_switch.do
Table 7	tables/tab_decongestion.tex	code/66_colombian_decongestion.do
Table 8	tables/tab_mech_capacity_wrap.tex	code/python/make_mech_tables.py
Table 9	tables/tab_fiscal.tex	code/64_fiscal.do
Figure A.1	figures/fig_a1b_permits_combined.png	code/14_fig_permits_combined.do
Figure A.2	figures/fig_a3_patients_panels.png	code/3_fig_a2_patients_pep_ppt.do
Figure A.3	figures/fig_a2b_permits_patients.png	code/15_fig_permits_patients.do
Figure A.4	figures/fig_a6_outpatient_ratio.png	code/17_fig_outpatient_ratio.do
Figure A.5	figures/fig_a7_annual_growth.png	code/18_fig_annual_growth.do
Figure A.6	figures/fig_ppt_by_regime.png	code/4_fig_a3_ppt_by_regime.do
Figure A.7	figures/fig_a8_care_type_groups.png	code/9_fig_care_type.do
Figure A.8	figures/fig_a11_consult_costs.png	code/13_fig_consult_costs.do
Figure B.1	figures/fig_a13_mortality_trends.png	code/20_fig_mortality_trends.do
Figure B.2	figures/fig_a14_outcome_trends.png	code/23_fig_outcome_trends.do
Table B.1	tables/tab_descriptives.tex	code/32_table_descriptives.do
Tables B.2–B.4	tables/tab_data_appendix.tex	code/65_data_appendix.do
Figure B (exposure map)	figures/fig_exposure_map.png	authored from data/clean exposure + DANE MGN boundaries (see paper notes). Generated manually using ArcGIS
Figure C.1	figures/fig_a20_event_studies_age.png	code/28_event_studies_age.do
Figure C.2	figures/fig_es_noncol.png	code/37_event_studies_nationality.do
Figure C.3	figures/fig_a17_trends_by_exposure.png	code/25_fig_trends_by_exposure.do
Figure C.4	figures/fig_a18_age_mortality_by_exposure.png	code/26_fig_age_mortality_by_exposure.do
Figure C.5	figures/fig_i1_es_outcomes_col.png	code/54_es_nationality_split.do
Figure C.6	figures/fig_i2_es_outcomes_ven.png	code/54_es_nationality_split.do
Figure C.7	figures/fig_i3_es_ages_col.png	code/54_es_nationality_split.do
Figure C.8	figures/fig_i4_es_ages_ven.png	code/54_es_nationality_split.do
Table C.1	tables/tab_descriptives_age.tex	code/32_table_descriptives.do
Table C.2	tables/tab_did_age.tex	code/40_table_did_age.do
Table C.3	tables/tab_did_mortality_ven.tex	code/31_tables_did.do
Figure D.1	figures/fig_h1_covid_clean.png	code/53_mortality_defs.do
Figure D.2	figures/fig_h2_clean_trends.png	code/53_mortality_defs.do
Figure D.3	figures/fig_h3_age_clean_trends.png	code/53_mortality_defs.do
Table D.1	tables/tab_mortclean_desc.tex	code/53_mortality_defs.do

Exhibit	Output file	Generating script
Table D.2	tables/tab_covid_excluded.tex	code/53b_covid_excluded_did.do
Table D.3	tables/tab_covid_nonparametric.tex	code/53c_covid_nonparametric.do
Figure E.1	figures/fig_e3_nightlights_es.png	code/46_nightlights.do
Figure E.2	figures/fig_h4_decongestion_es.png	code/66_colombian_decongestion.do
Figure E.3	figures/fig_g1_capacity_dynamics.png	code/50_mech_evidence.do
Figure E.4	figures/fig_g2_mediation.png	code/50_mech_evidence.do
Figure E.5	figures/fig_g3_workforce_es.png	code/50_mech_evidence.do
Figure E.6	figures/fig_g4_pyramid.png	code/51_pyramid.do
Table E.1	tables/tab_mech_homicide_wrap.tex	code/python/make_mech_tables.py
Table E.2	tables/tab_causes_wrap.tex	code/48_table_causes.do
Table E.3	tables/tab_causecap_wrap.tex	code/52_mech_evidence2.do
Table E (migrant ER composition)	tables/tab_mig_er_comp.tex	code/62b_migrant_er_composition.do (input slimmed by code/python/slim_disease_panel.py)
Figure F.1	figures/fig_f1_births_es.png	code/45_births.do
Table F.1	tables/tab_births_wrap.tex	code/45_births.do

5. Requirements

Software

- Stata 16 or later. The master script auto-installs the required user-written commands into a project-local code/ado folder (so versions travel with the package): require, ftools, reghdfe, estout, binscatter, distinct, and, for the robustness table, csdid, drdid, did_imputation, did_multiplegt_dyn, coefplot. These packages are also shipped pre-installed in code/ado. The Rambachan–Roth honest-DiD sensitivity (34_robustness.do and 34b_honest_weighted.do) additionally uses honestdid, which is not on SSC; main.do installs it from its GitHub repository. Both calls are wrapped in capture noisily, so the package still runs to completion if that install is unavailable offline — only the honest-DiD rows/statistics are skipped in that case.
- Python (tested with 3.11).
- The three table-assembly scripts invoked by main.do use the standard library only (csv, os, math).
- The two optional birth/population preparation scripts (only needed to regenerate the provided intermediates in Section 2.3) additionally require pandas, msoffcrypto-tool, openpyxl, and xlrd.

Reproducibility of stochastic results

- The robustness table's heterogeneity-robust estimators bootstrap their standard errors. main.do sets set seed 90210 (and 34_robustness.do resets the same seed before the estimator loop) so those columns reproduce exactly.

System

- Run time: about 20 minutes end-to-end.

6. Folder Structure

```

package/
├── code/
│   ├── 51 Stata do-files + Python builders + ado/ (installed at run time)
│   └── main.do master script (change one path line, then run)
├── data/
│   └── inputs (Section 2); clean/ holds regenerated intermediates
├── outputs/tables
│   └── regenerated table fragments (.tex)
└── outputs/figures
    └── regenerated figures (.png)

```