Replication Guide: TRIPS, Pharmaceutical Patents, and Generic Competition in India

This guide provides code for replication of the analysis for "TRIPS, Pharmaceutical Patents, and Generic Competition in India."

This README is located in `Analysis_Replication/`. Only code and non-confidential data files are included in the public replication package. Confidential data must be obtained separately:

- **IQVIA Ark Patent Intelligence**: https://www.iqvia.com/solutions/industry-segments/generics/ark-patent-intelligence
- **IQVIA MIDAS**: https://www.iqvia.com/solutions/commercialization/data-and-information-management/midas

---

Directory Structure

```
replication_package/
|-- Code_Replication/          # All code files (00-18)
|-- NonConfidentialData/       # Public data (included)
|-- ConfidentialData/          # IQVIA, MIDAS (not included)
|-- processed/                 # Output datasets
|-- tmp/                       # Temporary files
+-- Analysis_Replication/      # Output figures and tables
```

Quick Start

```
cd "[path to replication_package]/Code_Replication"
do "00_run_all.do"
```

Prerequisites

Data Licenses

The following proprietary data must be licensed separately:

- **IQVIA Ark Patent Intelligence** — global patent landscape data
- **IQVIA MIDAS** — pharmaceutical market data for India and U.S.

Software

- **Stata 17.0** (required)
- **Python 3.x** (for UMLS mapping scripts)

Stata Packages

Install via SSC before running:

```
ssc install estout
ssc install sutex
ssc install strip
```

Python Packages

```
pip install pandas requests python-dateutil openpyxl
```

API Access

- UMLS API key (set as environment variable `UMLS_API_KEY`)

Data Pipeline Overview

The master script `00_run_all.do` executes files 01–18 in sequence.

Phase 1: Public Data Preparation

- `01_fda_data_preliminaries.do` — Extracts FDA NMEs from Drugs@FDA (1995–2017)
- `02_patent_data_preliminaries.do` — Processes Orange Book patent data
- `03_create_input_for_iqvia_patent_request.do` — Creates IQVIA patent request file

Phase 2: UMLS Mapping (Python)

Run separately (require API access):

- `04_map_arkname_to_umls.py` — Maps IQVIA ARK names to UMLS
- `05_map_midas_to_umls.py` — Maps MIDAS molecules to UMLS
- `06_map_applno_to_umls.py` — Maps FDA applnos to UMLS
- `07_map_gbd_to_umls.py` — Maps GBD diseases to drugs via UMLS

Phase 3: Disease Mapping

- `08_match_icd_to_gbd.do` — Maps ICD-10 codes to GBD causes

Phase 4: Drug Characteristics

- `09_create_drug_vars.do` — Creates drug characteristics dataset

Phase 5: Patent Data

- `10a_create_indian_patent_vars_applno_level.do` — Patent status at applno level (Figures 1–2)
- `10b_create_indian_patent_vars_midas.do` — Patent status at MIDAS level (regressions)
- `11_create_indian_patent_expiration.do` — Indian patent expiration dates

Phase 6: Market Data

- `12_create_competition_in_atc3.py` — Competition measures, India (Python)
- `13_create_competition_in_US.py` — Competition measures, U.S. (Python)
- `14_get_MIDAS_information.do` — Processes MIDAS India data
- `15_get_MIDAS_information_US.do` — Processes MIDAS U.S. data

Phase 7: Final Dataset

- `16_create_indian_market_vars.do` — Creates `dataset_for_analysis.dta`

Phase 8: Analysis & Output

- `17_make_exhibits.do` — Generates figures and tables
- `18_cr_datadiagram.do` — Generates sample flow diagram

Output Files

Figures (in `Analysis_Replication/figures/`)

- `f1ab.png` — Figure 1: U.S. Approval Year and PPPYear
- `f2ab.png` — Figure 2: Share with Indian Applications/Patents
- `f3_combined.png` — Figure 3: Independent Entry (India vs U.S.)
- `sample_flow_diagram.html` — Data overview flow diagram

Tables (in `Analysis_Replication/tables/`)

- `xtab_firms.tex` — Table 1: Pre/post-1995 comparison
- `descriptive_statistics_exact.tex` — Table A.1
- `ols_live_exact_*.tex` — Tables A.2–A.4
- `descriptive_statistics_all.tex` — Table A.5
- `ols_live_all_*.tex` — Tables A.6–A.8
- `ols_pppyear_*.tex` — Tables A.9–A.10

Contact

For questions about the replication package, contact the authors.