

# Data Analysis README

## Project Overview

This note aims to address this information gap by presenting stylized facts on the global health workforce.

## Overview of the reproducibility check folder:

We have reproduced the graphs in the note using both Stata and Excel. The “Scripts” folder has the dofile used to create graphs on Stata. The graphs from Stata will be stored in the figures folder within the output folder (please check the img sub-folder for png files of the graphs) once the “dofile” is run. The graphs created on Excel are in the Excel file in the “Reproducibility in excel” folder.

The final EFI note is in the Reports folder.

## Reproducibility in STATA: Files and Folders

Data: This folder contains all the input and output data files used in the analysis.

Within the Data folder, there are two sub-folders: *input and output*.

### input data sources:

- GDP Per capita constant 2015 USD from World Development Indicators (WDI) for 2000-2022. Data file name is “GDP Per Capita 2015 USD.xls”. Data is publicly available at Source: <https://data.worldbank.org/indicator/NY.GDP.PCAP.KD?end=2022&start=2001>.
- Most recent Worldwide Bureaucracy Indicators (WWBI). Data file name is “WWBIEXCEL.xlsx”. The Worldwide Bureaucracy Indicators (WWBI) are a unique cross-national dataset on public sector employment and wages. The indicators, constructed from nationally representative household surveys, present a micro-founded picture of public sector labor markets across the world. The WWBI are based on 53 million unique survey observations and consist of 112,919 estimations across 192 indicators for 202 countries and territories between 2000 and 2018. All the do files are that prepare the indicators are on GitHub: <https://github.com/worldbank/Worldwide-Bureaucracy-Indicators>.
- World Bank regional classifications of countries from the WDI. Data file name is ‘CLASS.xlsx’. Data is publicly available at Source: <https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups>.
- Number of medical workers and medical workers per capita is based on unpublished WWBI data. This was obtained directly from the authors of the WWBI and cannot be made public. This dataset is used only for Figure3. Data file name is “fig3\_medicalperpopulation.dta”, and in the excel sheet (for figure 3) in the “reproducibility in excel” folder.
- Public sector wage premium for healthcare workers compared to public administration workers based on unpublished WWBI data obtained directly from the authors of the WWBI, and cannot be made public. This dataset is used only for Figure11. Data file name is “fig11\_wagepremium.dta” on Stata.

### output folder also has two sub-folders.

- data: This folder has the below data files for the analysis:
  - Final merged dataset for analysis. Name of the file is “finaldata-forfigures.dta”.
  - Global averages by indicator: Name of the file is “global\_averages.dta”
  - Regional averages by indicator: Name of the file is “regional\_averages.dta”
- figures: This folder has the figures prepared by the do file. It has a subfolder named “img”, where all the figures are being saved in png formatting.

### Scripts:

- This folder contains the do file used for data cleaning, processing, analysis, and visualization. Name of the file is “dofile.do”.
- The do-file runs the entire code that merges datasets, saves the intermediate ones in the input data folder, and final 3 datasets for analysis in the output data folder.
- It also creates the graphs that are in the note, and saves them in the figures folder in the output folder.

### Notes:

- Please note that the final figures in the report are from Excel, and some of those (mentioned below) were created using Data Wrapper.
- The reproducibility on Excel is explained below.
- Figures 1, 2, 5a, 5b, 6, 7, 8, 9, 11, 12a, 12b, 13 are displayed as percentages in the report but proportions in the figures from Stata.

### Reproducibility in EXCEL: Sheets

- Sheet 1 “WWBI original”: Contains the raw WWBI data.
- Sheet 2 “Full data\_latestavailableyear”: Presents the most recent data available for each indicator by country, merged with GDP per capita and regional identifiers (this is done on Stata and then exported to excel). Refer to the Stata do file for detailed steps on data preparation. This sheet is identical to “finaldata-forfigures.dta”.
- Sheet 3 “Global average (collapsed)”: Data is coming from “global\_averages.dta” file. It is Sheet 2 collapsed by indicator on Stata, providing global averages per indicator.
- Sheet 4 “Regional average (collapsed)”: Data is coming from “regional\_averages.dta” file. It is Sheet 2 collapsed by indicator and regions on Stata, providing regional averages per indicator.
- Sheet 5 “fig1\_industry\_share\_hlth”: Displays “Figure 1: Globally, over two-thirds of the paid healthcare workforce are employed in the public sector.”, utilizing data from Sheet 4.
- Sheet 6 “fig2\_health\_publicpaid”: Features “Figure 2: The share of the health workforce in public sector employment”, using data from Sheet 2.
- Sheet 7 “fig3\_med\_pc”: Showcases “Figure 3: Country income and supply of medical workers are positively correlated”. This sheet employs data from “fig3\_medicalperpopulation.dta”, which includes the number of medical workers and medical workers per capita based on unpublished WWBI data obtained directly from the authors of the WWBI.
- Sheet 8 “fig4\_mean\_age”: Depicts “Figure 4: Mean age of paid employees by industry, health”, using data from Sheet 4. The final figure in the report was created using Data Wrapper.

- Sheet 8 “fig5\_tertiary\_publicvsprivate”: Illustrates “Figure 5: Many countries face shortages of skilled healthcare workers.”, utilizing data from Sheet 4. The final figure in the report was created using Data Wrapper.
- Sheet 9 “fig6\_female\_emp\_health”: Exhibits “Figure 6: Globally, most public healthcare workers are women.”, based on data from Sheet 4.
- Sheet 10 “figure7\_femaleoccupations”: Presents “Figure 7: Women are only 35 percent of all managers in the public sector”, using data from Sheet 4.
- Sheet 11 “fig8\_wageprem”: Features “Figure 8: Public sector health workers receive a wage premium”, sourced from Sheet 3.
- Sheet 12 “Fig9\_sizeofwagepremium”: Contains “Figure 9: The size of the public sector wage premium varies significantly across countries.”, utilizing data from Sheet 2.
- Sheet 13 “Fig10\_crosscomparison”: Includes “Figure 10: Wages of hospital doctors and nurses relative to the global median by region”, sourced from Sheet 4. The final figure in the report was created using Data Wrapper.
- Sheet 14 “fig11\_regression-extra”: Presents “Figure 11: Public sector health workers are paid less compared to public”, utilizing data from “fig11\_wagepremium.dta” based on unpublished WWBI data obtained directly from the authors of the WWBI.
- Sheet 15 “Fig12wagepremiumfemale”: Showcases “Figure 12: In most regions, women tend to earn a higher wage when employed in the public sector healthcare as opposed to the private sector.”, using data from Sheet 4.
- Sheet 16 “fig13\_genderwage\_health”: Displays “Figure 13: Female health workers receive a wage penalty compared to their male counterparts.”, utilizing data from Sheet 4.