Healthcare Rationing in Public Insurance Programs: Evidence from Medicaid README

Timothy J. Layton* Nicole Maestas† Daniel Prinz‡ Boris Vabson§

October 20, 2021

1 Raw Data

The directory /raw/ contains the following raw datasets.

1.1 Medicaid Analytic eXtract

We obtained the Medicaid Analytic eXtract (MAX) through Data Use Agreement (DUA) 25543 from the Centers for Medicare and Medicaid Services (CMS) through the National Bureau of Economic Research (NBER). For more information on how to obtain the MAX data and other research identifiable files (RIF), see the website of the Research Data Assistance Center (ResDAC):
https://www.resdac.org/research-identifiable-files-rif-requests.

Codebook: Codebooks for the MAX files are available online:

- https://resdac.org/cms-data/files/max-ip/data-documentation
- https://resdac.org/cms-data/files/max-ot/data-documentation
- https://resdac.org/cms-data/files/max-rx/data-documentation
- https://resdac.org/cms-data/files/max-ps/data-documentation

Accessibility: The data are confidential, but may be obtained with Data Use Agreements with the Centers of Medicaid and Medicare Services (CMS). For more information on how to obtain the Medicare data and other research identifiable files (RIF), see the website of the Research Data Assistance Center (ResDAC):

*Harvard University and NBER. Email: layton@hcp.med.harvard.edu
†Harvard University and NBER. Email: maestas@hcp.med.harvard.edu
‡Harvard University. Email: dprinz@g.harvard.edu
§Harvard University. Email: vabson@hcp.med.harvard.edu

https://www.resdac.org/research-identifiable-files-rif-requests. The DUA requires secure storage and data management protocols, as well as IRB approval. The cost is approximately $2,000/file-year. Graduate students can access the data for free. The data must be loaded under /raw/max/data/.

1.2 Clinical Classification Software Data

The Clinical Classifications Software (CCS) is a classification developed as part of the Healthcare Cost and Utilization Project (HCUP) by the Agency for Healthcare Research and Quality (AHRQ). It groups diagnosis codes into clinically meaningful categories. For our analysis, we used the highest level of aggregation with 18 groups and present results for the 10 most common categories. The CCS classification is available online at https://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp. This data can be found under /raw/ccs/data/.

1.3 Therapeutic Class Data

We used the RED BOOK classification associated with Truven Health's MarketScan data to classify prescription drugs into therapeutic classes. This is a proprietary dataset obtained through the NBER. It can be purchased at this website: https://www.ibm.com/products/micromedex-red-book. The data must be loaded /raw/marketscan/data/.

Accessibility: The RED BOOK classifications associated with Truven Health's MartketScan data are proprietary and confidential. These data provide drug characteristics and classes by National Drug Codes (NDCs). These data was obtained through the NBER and can be purchased at: https://www.ibm.com/products/micromedex-red-book. The cost of this data is approximately $3,000.

Codebook: The codebook for the RED BOOK data is provided under /raw/marketscan/docs/

1.4 Texas County Data

We used geographic data on counties in our analysis that can be found under /raw/counties/data/. The following datasets are located in this directory:

- PHR_MSA_County_masterlist.xlsx
  - Contents: This file contains basic information on Texas counties.
  - Source: Texas Department of State Health Services
  - Downloaded 3/18/2017 from https://www.dshs.texas.gov/chs/info/info_txco.shtm.
- county_adjacency.txt
  - Contents: The county adjacency file lists each county, or county equivalent, and which county, or counties, are neighboring. The file includes all 50 states, the District of Columbia, Puerto Rico and the Island Areas (American Samoa, the Commonwealth of the Northern Mariana Islands, Guam, and the U.S. Virgin Islands).
  - Accessibility: This is a public use data file from the US Census Bureau.

  - Source: US Census Bureau
  - Downloaded 3/18/2017 from https://www.census.gov/geographies/reference-fi 2010/geo/county-adjacency.html.

- cb_2015_us_county_500k.dbf, cb_2015_us_county_500k.shp, cb_2015_us_county_500k.shx
  - Contents: These files are the 500k or 1:500,000 ESRI shapefiles for US counties. The cartographic boundary files are simplified representations of selected geographic areas from the Census Bureau's MAF/TIGER geographic database. These boundary files are specifically designed for small scale thematic mapping.
  - Accessibility: This is a public use data file from the US Census Bureau.
  - Source: US Census Bureau
  - Downloaded 3/18/2017 from https://www.census.gov/geographies/mapping-file time-series/geo/carto-boundary-file.2015.html.

- map_data_database.dta, map_data_coordinates.dta
  - Contents: These files are the database and coordinates file for plotting the county maps using spmap in Stata.
  - Source: Created from cb_2015_us_county_500k.dbf, cb_2015_us_county_500k.shp, cb_2015_us_county_500k.shx using the command

    ```
    shp2dta using cb_2015_us_county_500k, database(map_data_database)
    coordinates(map_data_coordinates) genid(id) replace
    ```

    in Stata.

1.5 Texas Zipcode Data

We used geographic data on zipcodes in our analysis that can be found under /raw/zipcodes/data/. The following datasets are located in this directory:

- cb_2013_us_zcta510_500k.dbf, cb_2013_us_zcta510_500k.prj, cb_2013_us_zcta510_500k.shp, cb_2013_us_zcta510_500k.shp.iso.xml, cb_2013_us_zcta510_500k.shp.xml,

  cb_2013_us_zcta510_500k.shx, zcta510_500k.ea.iso.xml
  - Contents: These files are the 500k or 1:500,000 ESRI shapefiles for US zipcodes. The cartographic boundary files are simplified representations of selected geographic areas from the Census Bureau's MAF/TIGER geographic database. These boundary files are specifically designed for small scale thematic mapping.
  - Accessibility: This is a public use data file from the US Census Bureau.
  - Source: US Census Bureau
  - Downloaded 11/13/2018 from https://www.census.gov/geographies/mapping-fil time-series/geo/carto-boundary-file.2013.html

- sf12010zcta5distance25miles.dta

  - Contents: Contains the distance between zipcodes that are within 25 miles.
  - Source: NBER
  - Downloaded 11/13/2018 from
    http://www.nber.org/data/zip-code-distance-database.html

- map_data_database.dta, map_data_coordinates.dta
  - Contents: These files are the database and coordinates file for plotting the zipcode maps using spmap in Stata.
  - Source: Created from cb_2013_us_zcta510_500k.dbf, cb_2013_us_zcta510_500k.shp, cb_2013_us_zcta510_500k.shx using the command
    `shp2dta using cb_2013_us_zcta510_500k, database(map_data_database) coordinates(map_data_coordinates) genid(id) replace` in Stata

1.6 State Drug Utilization Data

We used annual data on prescription drug use by drug to compare Arkansas and Texas when in 2012 they carved prescription drugs into Medicaid managed care in Texas. The Medicaid State Drug Utilization Data is freely available online from https://www.medicaid.gov/medicaid/prescription-drugs/state-drug-utilization-data/index.html. The data must be loaded under /raw/state_drug_utilization_data/data/.

2 External Analyses

We provided the Social Security Administration with code to estimate the impact of the STAR+Plus program on employment, mortality, and suspensions using the data. The SAS code we provided them with and the output generated can be found under /import/ssa_analysis/, along with code that imports their output into Stata.

3 Data Construction

To create the datasets that are used in our analyses to create the figures and tables, the directories under /derived/ should be the run in the following order.

1. Counties
   (a) /derived/counties_mmc/
2. Zipcodes
   (a) /derived/zipcodes/
3. Sample creation
   (a) /derived/sample_mmc/

4. Price imputation and price adjustment
    (a) /derived/price_imputation/
    (b) /derived/price_adjustment/
5. Providers
    (a) /derived/provider_list/
6. Utilization
    (a) /derived/utilization_mmc/
    (b) /derived/utilization_rx_mmc/
    (c) /derived/utilization_rx_detail_mmc/
    (d) /derived/utilization_ps_mmc/
    (e) /derived/utilization_op_rx_mmc/
7. Elixhauser Score
    (a) /derived/elixhauser_mmc/
8. Compile main analytical dataset
    (a) /derived/compiled_data_mmc/
    (b) /derived/compiled_data_quarterly_mmc/
    (c) /derived/compiled_data_annual_mmc/
    (d) /derived/compiled_data_noncontiguous_mmc/
    (e) /derived/compiled_data_annual_noncontiguous_mmc/
9. State drug utilization data
    (a) /derived/state_drug_utilization_data/

To run each of the directories change the current directory to /code/ within that directory.

4 Figures and Tables

The figures and tables in the paper can be generated by running /analysis/figures_and_tables/code/figures_and_tables.do.

5 Additional Programs and Settings

The directory /lib/ contains the file globals_mmc.do which is called at the beginning of each program. It contains various parameter settings and variable labels that are used across different do-files. The directory /lib/ado/ contains programs that are called across do-files or called multiple times in /figures_and_tables/.

6 Replication Requirements

This replication kit requires the following computation resources

- Stata 15 or higher with the following packages installed (these ado files are included in the directory "`/lib/`"):
  - `reghdfe`
  - `coefplot`
  - `estout`
  - `spmap`
  - `elixhauser`
- Computational requirements:
  - Cluster size: 300 G and Equivalent RAM
  - Disk Size: 1,000 G
  - Memory Size: 200 G
- Time requirement: 4 days