README for “The Corporatization of Independent Hospitals” by Elena Andreyeva, Atul Gupta, Catherine Ishitani, Malgorzata Sylwestrzak, and Benjamin Ukert

January 2024

Accessing proprietary data

In this replication package, we include the STATA code and public data used to produce the final models, figures, and tables in the paper. Since our main data sources are proprietary, we cannot share these data in the replication package. However, the final results can be replicated using the enclosed code by any researcher with access to the data.

Our analysis primarily relies on four proprietary data sources, which we describe in Section 3 of the paper and Sections A1-3 of the online Appendix.

1. Elevance Health inpatient claims
2. Traditional Medicare inpatient claims
3. New York inpatient discharge data
4. American Hospital Association (AHA) survey data

Elevance Health has previously worked with external researchers in formal collaborations under data use agreements. Traditional Medicare claims may be purchased from the Centers for Medicare & Medicaid Services (CMS) under a data use agreement. Some research institutions already have access to Medicare claims under a DUA (e.g., NBER). Researchers at such institutions can apply to CMS to reuse the data at a greatly reduced fee. New York inpatient discharge data may be purchased from the Agency for Healthcare Research and Quality (AHRQ). Finally, AHA data may be purchased from the AHA or accessed through the NBER.

Replication steps

We summarize the steps required to replicate our results using the public and proprietary data sources.

1. Download the replication package and the STATA software dependencies described in the last section of the README file. Unzip the public_data and public_data/hcris_data/csv folders.
2. Run the following do files in order:
    a. **0_main_public.** Processes the public data sources. Transfer the output files from the public_data output folders to an Elevance server.
    b. **1_main_elevance** on an Elevance server. Creates and analyzes the final Elevance data, generating figures and tables. Transfer the (non-sensitive) .dta files from the output folder to the Medicare and NY discharge servers.
    c. **2_main_medicare** on a server with Medicare data. Creates and analyzes the Medicare data.
    d. **3_main_NY** on a server with NY discharge data. Creates and analyzes the final discharge data.

    The do files are described in further detail in the next section.

Do files

0_main_public calls the do files in the public_code folder:

- **import_hcris:** Imports the raw hospital cost report (HCRIS) data into STATA.
- **process_hcris**: Processes the imported HCRIS data, extracting relevant FTE and financial variables and calendarizing the data.
- **create_acs_controls:** Processes the raw ACS data.
- **create_hcahps:** Imports the raw patient experience (HCAHPS) data into STATA.

1_main_elevance calls the do files in the elevance_code folder:

- These files assume the construction of episode files from Elevance, which include spending summary and readmission variables. The episode files have been merged to AHA IDs and mergers and saved at the episode and hospital-year levels. These processing steps are described in Appendix A1-A4.
- **create_hcris_vars:** Create FTE and margin variables from the processed HCRIS data.
- **create_regdata_FULL:** Create the final hospital-wide regression files: select the analysis sample and normalize dependent variables for prices and from the AHA, HCRIS, and HCAHPS.
- **match:** Match treated to control hospitals using coarsened exact matching.
- **create_deals:** Using the merger data, classify and summarize corporatization and non-corporatization deals. Generates *Table 2, Figure 2(a).*
- **create_regdata_QUALITY:** Create the final cardiac regression files: select the analysis sample and standardize readmission dependent variables.
- **analysis_aha:** Generate maps of the hospital mergers and cross-sectional scatterplots of the episode data. Creates *Figures 1, 2(b), A1(b), A2.*
- **analysis_FULL:** Run main DD regressions for Elevance prices, inputs, patient experience, and volumes. Generates *Tables 3, 5, 6, A4, A6, A8, A9; Figures 2(a), 3, 5, A4, A5, A6.*
    - Run heterogeneity, robustness, and patient mix regressions for *Tables 3, 4, 7, A3, A5, A6, A7, A10*; *Figure A3(a).*
    - Calculate standardized differences between matched and unmatched treated and control groups, *Table A2.*
    - Test corporatization vs non-corporatization coefficients for *Figure 4.*
    - Generate summary statistics for *Tables 1, A1.*
- **analysis_QUALITY:** Run DD regressions for Elevance readmissions. Generates *Tables 6, 7, A3, A7, A9, A10; Figures 7, A5(a), A7.*
- **extensive:** Run DD regressions for cardiac and delivery extensive margins. Generates *Tables 5, A8; Figure A3(b).*
- **create_scatterdata:** Run bootstrapped DD regressions for prices, inputs, and readmissions for each individual treated hospital.
- **analysis_scatterdata:** Using the estimates from create_scatterdata, estimate passthrough rates. Generate scatterplots of readmission, labor, expenses, and passthrough effects, *Figures 6, A3(c,d), A5(c).*

2_main_medicare calls the do files in the medicare_code folder:

- **0_prog_master_vfinal**: Define a flag for a non-deferrable condition using ICD9/ICD10 codes.
- **1_build_inpatient_sample_vfinal:** Read in the claims data and selects the analysis sample (Elevance hospitals, non-deferrable conditions). Generate risk histories, readmissions, and mortality variables.
- **2_mort_reg_vfinal:** Run the main DD regressions for Medicare readmissions, mortality, and patient mix. Partially generate *Table 6(b); Table A3(d); Figures 7(d,e).*
- **3_build_descriptive_sample_vfinal:** Read in the claims data for 2012 and 2013. Generate readmissions and mortality variables for patients with non-deferrable conditions.
- **4_desc_summ_vfinal:** Generate summary statistics for *Table 1(d).*

3_main_NY creates and analyzes the final discharge data. It selects the analysis sample and runs the main DD regressions for NY discharge readmissions, in-hospital mortality, patient mix, and volume. Partially generates *Tables 6, A3, A4, A7, A9; Figures 7, A5, A7.*

Data Files (Public Data)

We use several publicly available data sources, described in Appendix A1.

1. **Healthcare Cost Report Information System (HCRIS).** Hospital cost reports are used to generate overhead and contract personnel variables and patient care operating margins. These were obtained from https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Cost-Reports.
2. **American Community Survey (ACS)**. ACS data are used to generate county-level control variables. These were obtained from Social Explorer (www.socialexplorer.com) and may also be obtained from NHGIS IPUMS (https://usa.ipums.org/usa/).
3. **Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS).** HCAHPS survey data are used to generate a composite patient experience score. These were obtained from https://data.cms.gov/provider-data/.
4. **BLS Consumer Price Index Detailed Expenditure Categories.** BLS data are used to generate Figure A1(a) and were obtained from https://www.bls.gov/data/home.htm.

public_data/hcris_data includes:

- **lookup.xlsx:** Contains the locations, types, and labels of variables in the HCRIS data.
- **csv.zip**: Contains raw HCRIS data in .csv format for all Medicare hospitals.

public_data/acs_data: Contains raw 1-year and 5-year ACS data in .dta format for all US counties.

public_data/hcahps_data: Contains raw HCAHPS data in .csv format for all surveyed hospitals.

public_data/bls_data: Contains raw BLS data in .csv format for all detailed expenditure categories.

Software Dependencies

The following external STATA packages were installed from SSC: gtools, egenmore, tabout, distinct, outreg2, reghdfe.