Apr-16-2020 Hang Yang

demographic Data analysis of physician fee schedule on insurance

# our data

Our vision is to set up anchor prices for physicians or patients’ reference on different medical procedure services around the demographic area in the United States. First, we have collected all data sources to compute the state level anchor procedure service prices for 51 states in the United States. Later on, we have discussed two special cases on state Indiana and Connecticut on the county level anchor prices.

In order to compute the anchor prices, we will need below data files:

## All Population data from United States Census Bureau

### As the latest demographic data is only released for year 2018. This sets timeline for our main targeting data files collected from Medicaid, Medicare, as well as other demographic files.

### Data is downloaded [here](https://www.census.gov/data/tables/time-series/demo/popest/2010s-counties-detail.html#par_textimage_1383669527).

### EDA on population

#### Data Structure: <Tabular> CSV file, make “STNAME”, “CTNAME”, “YEAR”, “AGRP” as primary/foreign keys, “STNAME” and “CTNAME” are nominal data type, while others are quantitative data type.

#### Granularity: Each row stands for a population statistics record in a county from a state within year 2007-2018. We can use Groupby + sum to get the total population of each state.

#### Scope: all states and counties statistics within year 2007-2018 in U.S.

#### Temporality: Time dependes on “YEAR” column which represents as 0-11 integers, where 0 represents 2007, 11 represents 2018. It’s too huge. We just select data for year 2018 in the end.

#### Faithfulness: credited from United States Census Bureau

## Medicaid Enrollment Data

### State-Level

#### Data is downloaded here

#### EDA

### County-Level

#### Indiana

##### Data is downloaded here

##### EDA

#### Connecticut

##### Data is downloaded here

##### EDA

## Medicare Enrollment Data

### State and County all data file

#### Downloaded here

#### EDA

## Physician Fee schedule Data

### Medicaid Physician Fee Schedule

### Medicare Physician Fee Schedule

### Private Insurance PFS <Optional> Currently we assume it’s same as Medicare PFS.

### 

# data engineering and visualization

The Data we populated could be incomplete, noisy, dirty and invalid. Based on different data files and situations, we would use different data cleansing methods to correct the data frame.

## Noise data

### Invalid data: focus on the time span till 2018. For example, we will count price data which takes effect after Jan-1-2019 as invalid.

### Non statistics data: such as text messages/comments in the data file. Some of the files download has lots of comments and notes in the tail. We just ignore those rows.

## Missing data

### Blanks: usually we will ignore/filter out rows with blank value, such as “UNKNOWN” rows in Medicare enrollment data.

### “\*”, “nan”, “N/A”: if it’s population or enrollment data, usually we replace this data as 0. We will also check the context, such as in the Medicare enrollment data file, the totals are constructed by “FFS Beneficiaries” – Fee For Service Program and “MA Beneficiaries” – Medicare Advantage Program. Then we fill the “\*” in the total Medicare enrollment with “FFS Beneficiaries”.

## Merge data

### As our goal is to compute the demographic anchor PFS prices, we need to combine all population data file, Medicaid enrollment data file, Medicare enrollment data file, Medicaid PFS file, and Medicare PFS export report.

## Data Visualization

### 1, The estimated PFS distribution over all counties in per state with heatmap

### 2, The estimated PFS distribution over all states in U.S. with heatmap

### Together with Medicaid/Medicare enrollment rate, poverty rate

# output

## To compute the anchor procedure service price, we will take use the rates of Medicaid, Medicare, Private insurance enrollment rate in per state/county, and their physician fee schedule based on the HCPCT code/Procedure code.

## The Process is below:

## First, compute % of "STATE\_MEDICAID\_PRCENT" = Medicaid enrollment / state-level population, % of "STATE\_MEDICARE\_PRCENT" = # of Medicare beneficiaries / state-level population.

## Suppose % of private provider customers = 1 - % of "STATE\_MEDICAID\_PRCENT" - % of "STATE\_MEDICARE\_PRCENT

## Second, compute $ Anchored Physician Fee Amount for per procedure code = % Medicaid Rate \* Medicaid\_fee + % Medicare Rate \* Medicare\_fee + %Private Insurance Rate\* Private\_Insurance\_fee

## Here are some resulted tables.

## Table<1> Medcaid/Medicare/Private Insurance Enrollment Rate Per State

## A screenshot of a cell phone Description automatically generated

## State-Level PFS

### Indiana

## A screenshot of a cell phone Description automatically generated

### Connecticut

### A screenshot of a cell phone Description automatically generated

## County-Level PFS

### Indiana

## A screenshot of a social media post Description automatically generated

### Connecticut

## A screenshot of a cell phone Description automatically generated

# conlusion

## Through this project, we can get that different counties could have their own customized physician fee schedule based on demographic factor, poverty rate (Medicaid enrollment), aging factor (Medicare enrollment), etc. The physician fee schedule could present a waved trending line even in same sate.

## <Here could have some visualization is better?>

# Challenge and future

1, How to find valid, correct and uniformed formatted data is the key? But how could you get such data as each state has its own database or API?

2, How to combine these data together with other statistics data, like form CMS, Census bureau. Both data format and content is different, we need to figure out the relationship (primary+foreign key) in each data file.

3, How to collect the actual expenditure of the Medicaid, Medicare, and Private insurance in the real patient-visit cases? As our data is based on the CMS or each state’s release fee schedule, what is the real cost is unknown during the visit? These could be more valuable if possible.