

Israel Diego (UMID: israeldi)

Stats 504

November 12th, 2019

Medicare Services clustered by Geographical Structure

By working with Medicare data, we gain insight into the medical practice behavior for Medicare providers across the U.S. In particular, we determine whether providers are more inclined to use medical services in some geographical areas than others. Our initial approach for modeling the number of services provided is to use an additive quasi-Poisson regression model, which controls for provider type, U.S State, and the total number of services provided for each provider. We aim to study how controlling for various geographical structures affects the incidence of spatial clustering for each medical service.

Medicare Provider Utilization data comes from the Centers for Medicare and Medicaid Services website. The data contains information from 2017 on utilization, payment, and submitted charges organized by National Provider Identifier (NPI), Healthcare Common Procedure Coding System (HCPCS) code, and place of service.

The geographical structures of interest include: zip-code, census tract, county, and core-based statistical area (CBSA). **Table 1** gives us an idea of the degree to which each geographical structure encompasses providers within it, which is given by the number of clusters that each geographical structure has. We order the structures so that each structure is more encompassing than the one before it.

Table 1

Geographical Structure	Number of Clusters
Zip Code	244,243
Census Tract	16,247
County	2,969
CBSA	953

In order to model the mean-structure of the number of services provided for each HCPCS code, we take a GLM approach. We use quasi-Poisson regression for our GLM that is additive in its predictors, with the following variables. The dependent variable in our GLM is,

- **line_srvc_cnt**: the number of services provided.

The predictors in our model are,

- **provider_type**: categorical variable which identifies specific provider types such as nurse practitioners.
- **nppes_provider_state**: categorical variable which identifies the state where the medical service was provided
- **log_total_geo_cnt**: log of the total number of services provided for each NPI
- **log_total_npi_cnt**: log of the total number of services for each geographical structure. This variable is updated depending on the geographical structure we are clustering on.

Our aim is to gain insight into the correlation structure within geographical clusters. In this case, Generalized estimating equations (GEE) is appropriate. Since the number of clusters is large, we omit the geographical cluster variable. We measure the degree of spatial clustering by the intraclass correlation coefficient (ICC), which ranges from 0 to 1. A value of 0 would indicate that clusters in the residuals of our GLM are unrelated, while a value of 1 would indicate we have perfect clustering of the residuals.

We assess geographical clustering given by the ICC for each HCPCS code. We rank the top eight codes where the ICC was highest at the zip code level and compare to the ICCs of tract, county, and CBSA. Below in **Figure 1**, we display a bar chart of the ICCs for each code at each geographical structure. In **Table 2**, we can reference the medical services corresponding to the codes in **Figure 1**.

First, we observe that, compared to the other geographical structures, the ICC is always highest at the zip code level for each HCPCS code and also high at the census tract level. In the case of county and CBSA levels, it is not clear whether the ICC of county always dominates the ICC for CBSA or vice versa. Nonetheless, the ICC for county and CBSA is low for every code, with the maximum ICC over all HCPCS codes being 4% and 4.4% respectively. In other words, we cannot conclude that there is presence of unmeasured region-level covariates at the county and CBSA levels. Thus, the county and CBSA levels may be encompassing geographical areas too broadly for us to find any meaningful signs of spatial clustering.

For zip code and census tract, code (97140) resulted in the highest ICC, 0.49 and 0.18 respectively. At the zip code level we would conclude that there is a confounding factor at play and, to a lesser degree, at the census tract level. It is not always the case that high ICC for a code at the zip code level will result in a proportionately high level at the census tract level. This

can be seen for codes (98941) and (92014), where ICC for census tracts dips down and the ICC for zip code at these codes is higher. However, high ICC values for HCPCS codes at the zip code level can serve as a proxy for high ICC at census tract level, given that the top three codes for census tract managed to fall within the top eight codes of zip code ICCs.

One might wonder whether clustering at the zip code level would be preferred over the census tract level. We would argue that the zip code level may not be the most appropriate for clustering due to the fact that some zip codes consist of only one hospital and thus would give us perfect clustering in those cases. At the census tract level, we are still able to see signals of clustering similar to that of zip code level, but with a lower number of clusters. Thus the census tract geographical structure provides a good middle ground between zip code and county or CBSA levels.

In our analysis, we used ICC as a measure for spatial clustering within different geographical structures. We find that the degree of spatial clustering is highest for zip code and census tract geographical structures, and low for county and CBSA. However, we would recommend that clustering at the census tract level would provide more robust results, since each cluster would have a larger number of providers within it.

Figure 1

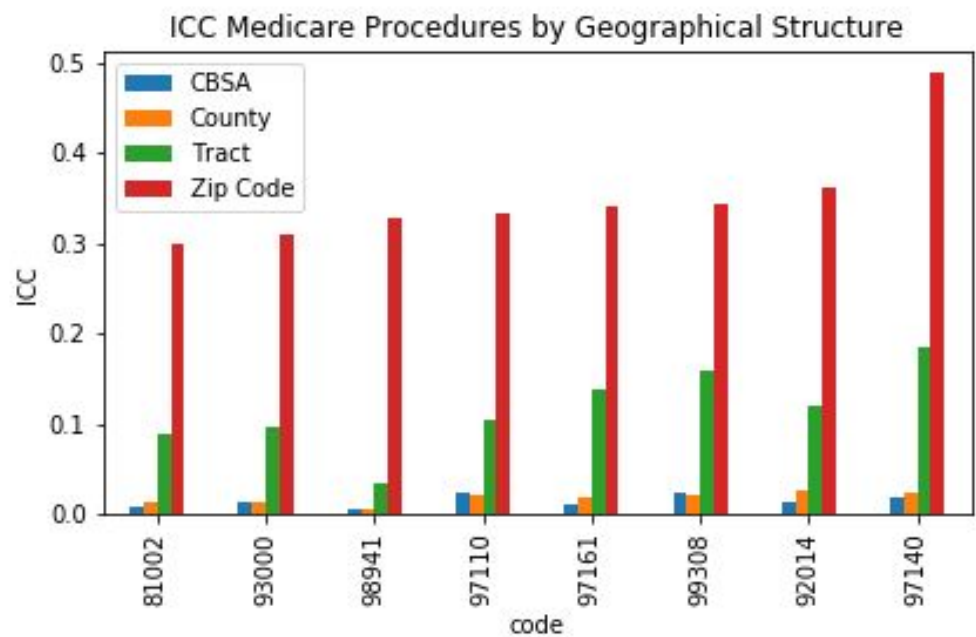


Table 2

HCPCS Codes	Description
81002	Urinalysis, manual test
93000	Routine EKG using at least 12 leads including interpretation and report
98941	Chiropractic manipulative treatment, 3 to 4 spinal regions
97110	Therapeutic exercise to develop strength, endurance, range of motion, and flexibility, each 15 minutes
97161	Evaluation of physical therapy, typically 20 minutes
99308	Subsequent nursing facility visit, typically 15 minutes per day
92014	Eye and medical examination for diagnosis and treatment, established patient, 1 or more visits
97140	Manual (physical) therapy techniques to 1 or more regions, each 15 minutes