Analyzing the Geospatial Rates of the Primary Care Physician Labor Supply in the Contiguous United States

By
Russ Frith
Advisor: Dr. Raid Amin
University of W. Florida
Capstone Project in Statistics
April, 2016
Abstract

• The national availability of primary care physicians remains steady or is declining in many areas of the nation.

• **Using publicly available data sets detailing county-based PCP supply, a detailed geostatistical analysis was done to characterize the adequacy of supply within the contiguous United States at the county level.**

• Statistical measures were computed that showed PCP labor distribution aligned along income distribution and ethnic composition of geographical regions.

• Those measures are reported using both choropleth maps depicting clusters of underserved populations and conventional statistical procedures.
Problem Statement

$H_0$ : The PCP county rates are randomly distributed

$H_A$ : The PCP county rates exhibit geographic clustering
Front Matter

- The Federal government agency responsible for monitoring and remediating the national PCP supply rate is the Health Resources and Services Administration (HRSA), National Center for Health Workforce Analysis division of the Health and Human Services Administration. HRSA is the authoritative source for geographic data sets that identify PCP rate.

- This agency compiles data on PCP supply in states, counties, and census tracts. It makes determinations if a geographic unit has a PCP labor shortage.

- HRSA has developed a protocol for PCP shortage determination based on localized count data.
Remediation grants and policies are formulated based on threshold counts typically tabulated at the county level.

The HRSA protocol for shortage calculations acknowledges that ethnic and low-income populations are at higher risk of being underserved and that such demographic cohorts are ubiquitous in nearly all counties.

Not taken into consideration through protocol however is the possibility that underserved populations are clustered.
Furthermore, PCP supply shortage calculations are not adjusted for other demographic factors such as disease prevalence, educational attainment, and age.

These protocols and assumptions have the potential for skewing the distribution of grants as well as the potential that qualified and eligible funding recipients are inadvertently excluded from such awards.

In addition, incentives formulated to rectify PCP rates may not be effective in that they do not address significant indicators of PCP rates.
Is this random? …or is there clustering?
Data Exploration
Data Sources

- County Health Rankings, http://www.countyhealthrankings.org/rankings/data
- Center for Disease Control Wonder http://wonder.cdc.gov/

Database:
http://rfrith.uaa.alaska.edu/ProseminarData.php
Methodology

- Factor-SaTScan™ Analysis
- Ordinary Least Squares and Geographic Weighted Regression
- Getis-Ord or “Hot-Spot” Analysis
- Poisson-SaTScan™ Analysis

http://www.satscan.org
Model Variables

- %5under: Percentage of county population which is under the age of 5.
- %Pov: Percentage of county population which has income below poverty level
- %I100KMore: Percentage of county population which has income of $100000 or more
- %LT50K: Percentage of county population which has income less than $50000
- %pubAsst: Percentage of county population which relies on public assistance
- %HSorLess: Percentage of county population which education attainment of high school or less
- %ColorMore: Percentage of county population which education attainment of a college degree or more
Model Variables

- PopDens: County population density
- %White: Percentage of county population which is Caucasian
- %Black: Percentage of county population which is black
- %Hispanic: Percentage of county population which is Hispanic
- %Unin: Percentage of county population which has no health insurance
- %Unemployed: Unemployment rate of a county
- %Med: Percentage of county population which is under Medicare coverage
Model Variables

• %Diabetes: Percentage of county population which has confirmed diabetes
• %Brain: Percentage of county population which has some form of brain cancer
• %YPLLRRat: Years of prolonged life lost ratio
• %Breast: Percentage of county population which has some form of breast cancer
• %Lung: Percentage of county population which has some form of lung cancer
• %Mort: Mortality rate of a county

http://rfrith.uaa.alaska.edu
Factor-SaTScan™ Analysis

\[ x_1 - \mu_1 = l_{11} F_1 + l_{12} F_2 + \cdots + l_{1m} F_m + \varepsilon_1 \]
\[ x_2 - \mu_2 = l_{21} F_1 + l_{22} F_2 + \cdots + l_{2m} F_m + \varepsilon_2 \]
\[ x_3 - \mu_3 = l_{31} F_1 + l_{32} F_2 + \cdots + l_{3m} F_m + \varepsilon_3 \]
\[ \vdots \]
\[ x_p - \mu_p = l_{p1} F_1 + l_{p2} F_2 + \cdots + l_{pm} F_m + \varepsilon_p \]

- \( \mu_i = \) mean of variable \( i \),
- \( \varepsilon_i = \) \( i \)-th specific factor,
- \( F_j = \) \( j \)-th common factor, and
- \( l_{ij} = \) loading of the \( i \)-th variable on the \( j \)-th variable.

\[ E(\varepsilon) = 0 \]
\[ \text{Cov}(\varepsilon, F) = 0 \]
\[ E(F) = 0 \]
\[ \text{Cov}(F) = I \]
Clusters of high Factor 1 scores: {Black, Unemployed, Poverty, Diabetes, Mortality}
Clusters of high Factor 2 scores: { Medicaid, Low income, Low educational attainment}

Question: What is the PCP distribution like in these clusters?
Cluster of high Factor 3 scores:
{High population density, high income, high education attainment, high brain cancer}

Question: What is the PCP distribution like in these cluster?
Cluster of high Factor 4 scores:
{High poverty, high public assistance, high mortality}

Question: What is the PCP distribution like in these clusters?
Cluster of high Factor 5 scores:
{Hispanic population, uninsured, high poverty}

Question: What is the PCP distribution like in these clusters?
The Getis-Ord local statistic is given as:

$$G_{i}^{*} = \frac{\sum_{j=1}^{n} w_{i,j} x_{j} - \bar{X} \sum_{j=1}^{n} w_{i,j}}{S \sqrt{\frac{n \sum_{j=1}^{n} w_{i,j}^2 - \left( \sum_{j=1}^{n} w_{i,j} \right)^2}{n-1}}}$$

(1)

where $x_{j}$ is the attribute value for feature $j$, $w_{i,j}$ is the spatial weight between feature $i$ and $j$, $n$ is equal to the total number of features and:

$$\bar{X} = \frac{n}{\sum_{j=1}^{n} x_{j}}$$

(2)

$$S = \sqrt{\frac{n}{\sum_{j=1}^{n} x_{j}^2} - \left( \bar{X} \right)^2}$$

(3)

The $G_{i}^{*}$ statistic is a Z-score so no further calculations are required.

The Gi* statistic returned for each county in the dataset is a Z score. For statistically significant positive Z scores, the larger the Z score is, the more intense the clustering of high values (hot spot). For statistically significant negative Z scores, the smaller the Z score is, the more intense the clustering of low values (cold spot).
The Gi* statistic returned for each county in the dataset is a Z score. For statistically significant positive Z scores, the larger the Z score is, the more intense the clustering of high values (hot spot). For statistically significant negative Z scores, the smaller the Z score is, the more intense the clustering of low values (cold spot).

Red: high z-scores or “hot spots”
Blue: low z-scores or “cool spots”
Clusters of high Factor 1 scores: { Black, Unemployed, Poverty, Diabetes, Mortality}

Question: What is the PCP distribution like in these clusters?
Clusters of high Factor 2 scores: { Medicaid, Low income, Low educational attainment}

http://rfrith.uaa.alaska.edu
Cluster of high Factor 5 scores:
{Hispanic population, uninsured, high poverty}
The clusters are regions of statistically low rates of PCP supply.

Clusters $\cap Z$ – scores

http://rfrith.uaa.alaska.edu
The clusters are regions of statistically high rates of PCP supply.

Clusters $\cap Z$ – scores

http://rfrith.uaa.alaska.edu
Conclusion:

Reject the null hypothesis in favor of the alternative hypothesis.

In 2012, the primary care physician rate at the county level was aligned along demographic strata.

Implications?
• Targeted grants
• Infrastructure remediation
• ACA Enforcement
• Detection protocol change?

http://rfrith.uaa.alaska.edu
Questions?

http://rfrith.uaa.alaska.edu