SAS2141-2018

Location Matters: Evidence from Spatial Econometric Analysis of Opioid Prescribing Rates

Manuel Figallo-Monge, Guohui Wu, and Jonathan McCopp, SAS Institute Inc.

ABSTRACT

A substantial number of Medicare patients are at risk of opioid abuse. A Centers for Medicare and Medicaid Services (CMS) analysis identified approximately 225,000 beneficiaries who received potentially unsafe opioid dosing. Among the beneficiaries with the highest number of opioid prescriptions filled in 2012, 23 opioid prescriptions were filled at an average cost of $3500 per beneficiary. A key to understanding beneficiary data effectively is to identify high-risk geographic clusters by applying location analytics based on historical data, demographic data, and health trends. Location analytics blends health data and socio-economic data with geographic data to reveal the location of opioid abuse among the Medicare population. Specifically, location analytics that includes spatial econometric modeling (like the SPATIALREG procedure) combined with SAS® Visual Analytics on SAS® Viya® is a powerful and easy-to-use solution to identify high-risk spatial clusters. It can better equip government agencies to effectively allocate resources where they are needed most to protect beneficiaries as well as boost the integrity of the government programs intended to help them. This paper shows how to apply location analytics to find improper prescriptions made by Medicare. It uses SAS® data management, data exploration, modeling, and reporting capabilities to identify patterns and relationships in data that address risks in Medicare and ultimately ensure a timely and adequate response.

INTRODUCTION

In 2015, the number of opioid-related deaths exceeded 33,000 for the first time.[1] Nearly half of these deaths involved prescription opioids. At first blush, statistics surrounding the opioid epidemic abound with numbers that shock and appear unbelievable.

In addition, the integrity of some United States Federal Government programs is under threat by the Epidemic since government programs, such as Medicare Part D, are at risk of being abused by illicit prescribing behaviors. Medicare Part D is the program that provides prescription drug benefits to U.S. citizens 65 and over. As noted in a recent OIG report, one in three Part D beneficiaries receive an Opioid.[1]

Furthermore, the Medicare Part D program has expanded significantly. From 2006 to 2015, total spending for Part D drugs increased by 167 percent, growing from $51.3 billion to $137 billion.[1] This is the amount that the Government, beneficiaries, and plan sponsors paid to pharmacies for Part D drugs. In 2015 alone, total spending increased by $15.6 billion, marking the third consecutive year that spending increases surpassed $10 billion.

The rapid expansion makes the integrity of the program vulnerable. One area that was particularly hard hit by the opioid epidemic is Appalachia, because much of the population have been involved in blue collar work where pain management, and medication are part of daily life. The Appalachian Region is a 205,000-square-mile region that follows the spine of the Appalachian Mountains from southern New York to northern Mississippi. It includes all of West Virginia and parts of 12 other states: Alabama, Georgia, Kentucky, Maryland, Mississippi, New York, North Carolina, Ohio, Pennsylvania, South Carolina, Tennessee, and Virginia.[19] Forty-two percent of the Region's population is rural, compared with 20 percent of the national population. Six percent of all Medicare prescription claims in the Appalachian Region are for opioids, compared to 5.3 percent for the United States as a whole.

The Region is also economically distressed. As previous research has indicated, poor areas such as those in Appalachia were allegedly targeted by pharmaceutical companies before the epidemic began.[2] Due to the impact of the epidemic, Appalachia has experienced significant negative impacts to local health, economy, and communities.

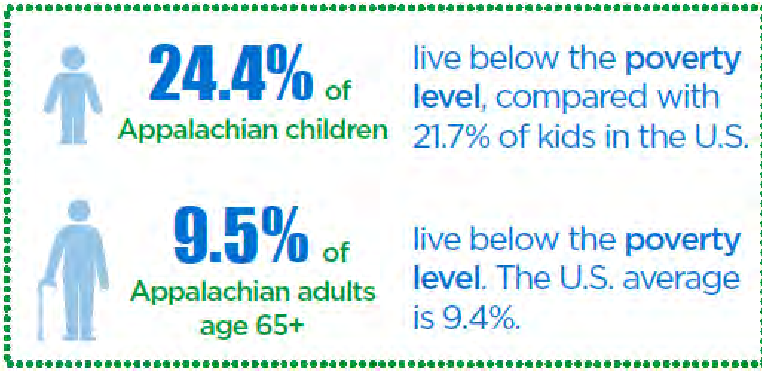

Figure 1. Poverty among Appalachian Children and the Medicare Adult Population.[19]

Figure 1 describes the poverty among the Appalachian children and the Medicare adult population. Since some of the communities have experienced a hollowing out of the young adult population due to opioids, their children are now being taken care of by grandparents. Figure 1 is an example of a link between the Medicare population and social determinates of health, such as poverty.

One of the goals of this analysis is to demonstrate through spatial econometrics how neighboring counties using Medicare in Appalachia are affected by the epidemic. It is well documented that certain counties (such as Mingo County in West Virginia) have been deluged by prescription drugs distributed by pharmaceutical companies.[3]

How does it put not only Medicare beneficiaries at risk but the Medicare program?

In other words, how do questionable prescribing patterns leading to Medicare fraud, waste, and abuse ultimately result from geographic or socio-economic determinants. Is the opioid crisis largely a result of the conditions in which people are born, grow, live, work and age?

BACKGROUND

To answer these questions a firm understanding of the use case, the analytics lifecycle, and SAS capabilities is required. Figure 2 shows these three topics which are covered in detail moving forward. Special attention is paid to spatial econometrics to gain insights that will help government agencies. Spatial econometrics is the field where spatial analysis and econometrics intersect.

Figure 2. Use Case, Analytics Lifecycle, and Capabilities.

Further, because the Medicare Part D data in this study is geo-referenced, the SPATIALREG procedure from the SAS Econometric and Time Series (ETS) package will be used and supplemented with visualizations from SAS Visual Analytics and SAS Visual Statistics.

USE CASE: A CLOSER LOOK

The opioid epidemic has impacted the United States Federal Government across many dimensions such as public health research, health care regulation, law enforcement, down to patient-provider encounters (See Appendix A). This paper will focus on health economics to understand opioid prescribing rates among the Medicare Part D population.

Econometric analysis provides an excellent means to study the functioning of the Medicare Part D health care system and "health affecting behaviors" such as opioid addiction. More broadly speaking, health economists can offer substantive expertise and methods for inferring key relationships within government data. Their expertise includes how health markets behave and how policy affects markets and market-related behavior.

The use case focuses on the Medicare (Centers for Medicare and Medicaid Services) Part D market, which is unique in many ways:

- The population is typically 65 and older
- It covered 43.6 million beneficiaries in 2016
- Commonly abused opioids accounted for over $4B in Part D spending

As expected, the volume and complexity of the data in this market has been a challenge for the health care and data science community. The data is complex, cross-sectional, and longitudinal. The challenges are addressed with the use of SAS capabilities that can help traverse the entire analytics lifecycle in one single platform as shown in Figure 2.

A health economics approach in understanding the opioid epidemic within the context of Medicare Part D necessarily involves mention of supply and demand. Figure 3 demonstrates the relationship between poverty estimates and Medicare variables that represent supply (Part D opioid prescribers) and demand (Part D opioid claims). The SAS Visual Analytics Correlation Matrix can be used to understand a variety of government variables used in supply side analysis (left) and demand-side analysis (right) as shown in Figure 3. In both examples, understanding poverty estimates is important since this presents a liability for Medicare Part D's low-income subsidies and cost sharing. In other words, Medicare low-income subsidies may be particularly vulnerable to abuse in areas with a high number of Part D prescribers or opioid claims.

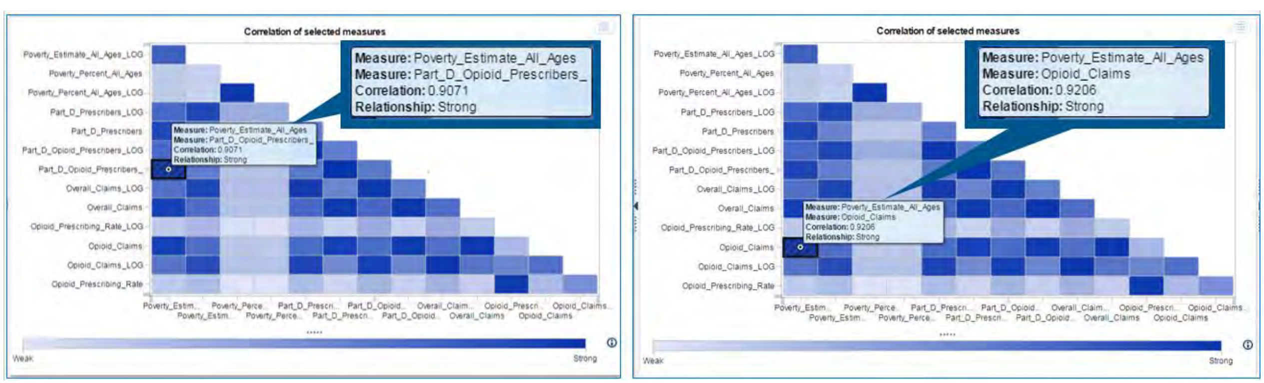

Figure 3. Using the SAS Visual Analytics Correlation Matrix to Understand Government Variables

Recent publications have shown that parts of the Appalachia region, which are especially poor, have been flooded with prescription drugs. Several notable examples exist in Mingo County where a glut (over-supply) could be especially pernicious in areas with a high number of prescribers involved with pill mills. Pill mills are places where doctors or unscrupulous providers hand out prescription drugs like candy. It is a term used primarily by local and state investigators to describe a doctor, clinic, or pharmacy that is prescribing or dispensing powerful opioids inappropriately or for non-medical reasons.

An over-supply of opioid prescriptions in a market or community is a risk to Medicare because it can increase the likelihood of questionable prescribing patterns and lead to the increased risk of fraud, waste, and abuse associated with Medicare reimbursements for prescription drugs.

Excess demand for opioid prescription drugs can also adversely impact the proper functioning of the Medicare system, especially when it, in the aggregate, is due to beneficiaries' illicit behaviors. Excess demand is likely an indication of Doctor Shopping and might be linked to increases in Medicare Part D opioid claims (as suggested earlier). As shown in Figure 4, Doctor Shopping occurs when an individual patient or beneficiary visits multiple providers or doctors to deceptively attain multiple prescriptions for controlled substances, such as opioids, and then visits multiple dispensers to acquire the opioids. This practice is not only against the law, but it also poses grave risks to those who engage in it, including legal penalties, addiction and overdose, not to mention the impact that this behavior has on their communities. Addiction and overdosing are especially noteworthy since the demand for opioids is inelastic –in other words, the demand for opioids will not change much if there is a change in price[18].

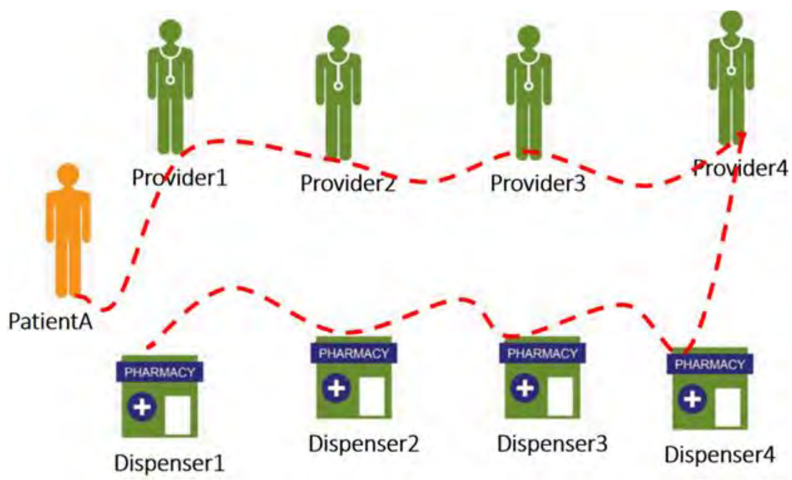

Figure 4. Doctor Shopping[1]

Against the backdrop of these dynamic market forces lie the social fabric of Appalachian communities, a fabric that has been weakened by the opioid crisis. Social determinants have a significant impact on health outcomes. Social determinants of health are "the structural determinants and conditions in which people are born, grow, live, work, and age."[4] They include factors like socioeconomic status, education, the physical environment, employment, and social support networks, as well as access to health care.

Social determinants not only help us understand the communities in which the Medicare population lives, but can help us to begin to identify geographic hotspots where health or policy interventions might be needed most. Hotspots can also lead investigators to areas where close examination of provider-patient interactions can be warranted.

This is also why predicting "hotspots" of high opioid prescribing rates where there is a possibly high supply and demand of opioids can be illuminating. To help gain this insight, SAS provides an analytics lifecycle as a framework to give analytic projects forward momentum.

ANALYTICS LIFECYCLE

The path forward in this research requires following the Analytics Lifecycle. For the purpose of producing a prototype, the analytics lifecycle consists of four phases:

- Data Management. In this phase, data is moved from one location to another using ETL (Extract, Transform, and Load) UI or User Interface components. For example, data can be extracted using API calls to a central repository, transformed long to wide for analytic processing, and loaded into memory for fast access to reports. APIs or Application Programming Interfaces are interfaces that enables communication between distributed applications.
- Data Explorations. This phase is the first step in data analysis and typically involves summarizing the main characteristics of a data set. It is commonly conducted using SAS Visual Analytics, but can also be done in more advanced statistical software, such as SAS Visual Statistics. Visual exploration helps the health economist understand what is in a data set and the characteristics of the data. These characteristics can include size or amount of data, completeness of the data, correctness of the data, possible relationships amongst data elements or variables in the data.
- Model Building. Model building is the process of developing a probabilistic model that best describes the relationship between the dependent and independent variables. The major issues are finding the proper form (linear or curvilinear) of the relationship and selecting which independent variables to include. Besides creating models, it's also important to test and compare models to choose the best fitting one.
- Deploying Models and Reports. The champion model from the previous step can be applied to a data set in a process called deploying or scoring. The data as a result is enriched with additional variables that result from applying the model. This enriched data set can be used to produce reports and even dashboards to broadcast results to an entire organization. As in the case of Medicare opioid prescribing rates, key policy decisions can be made and evaluated

The four phases are show in Figure 5 and ultimately transform data into insights.

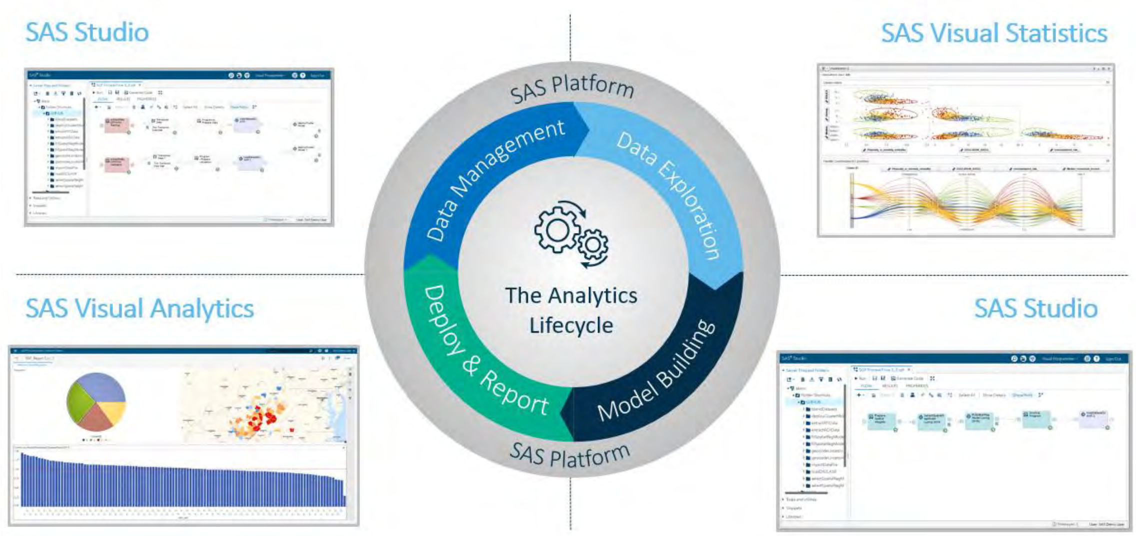

Figure 5. The Analytics Lifecycle

It is also useful for developing prototypes as we have done in the Medicare opioid research. Prototyping is an essential part of some analytic methodologies such as Agile that promote the concept of continuous improvement or incremental iterations. It's an essential step toward identifying and understanding project risks, gathering requirements, uncovering skills gaps, gaining technical understanding of the analytics problem, and mustering organizational support and resolving any conflicts. In Agile terminology, it allows projects to “fail fast”.

The content in this paper is meant to provide SAS customers enough material to build their own prototype and become more analytically mature using the three technologies in Figure 5: SAS Studio, which is a developmental web application for SAS that you access through your web browser; SAS Visual Statistics which offers a leading predictive analytics solution to explore data and build analytical models for a variety of use cases; and, finally, SAS Visual Analytics to allow users to publish reports very quickly to identify patterns, trends, and opportunities for further analysis through a variety of visualizations.

CAPABILITIES

As shown in this paper, an end-to-end solution was developed that followed the Analytics Lifecycle using SAS capabilities available on the SAS Platform. We covered all the steps of the analytics lifecycle in one integrated environment while emphasizing the following three attributes or characteristics of a quality analytics solution:

- **Ease-of-use.** The degree to which the analytics solution can be used by specified users to achieve quantified objectives with effectiveness, efficiency, and satisfaction all the while making analytics approachable.
- **Integrated and extensible.** Health care analytic solutions should take advantage of APIs that provide maximum flexibility for integration into custom processes and applications. SAS provides the platform to ensure that this is possible as well as bring together several sub-systems into one system so that the solution can deliver the overarching requirements and capabilities. In addition to linking different technologies into a coordinated whole, the SAS Platform takes future growth into consideration. This is the ability to extend the solution through the addition of new capabilities or the modification of existing capabilities. The central theme is to allow for change while minimizing impact to the existing solutions capabilities.
- **Insights.** This is achieved in two ways: through analytic models and visualizations. Models can use descriptive or predictive analytics techniques to gain knowledge from data. Graphs, charts, or maps can be used to visualize data to better understand trends and patterns in the data.

The SAS capabilities discussed are meant to demonstrate how the platform exemplifies each of these attributes.

Ease of Use

Drag-and-drop capabilities in a web interface and reusable software code driven by input variables to a coding interface make the SAS solution for Medicare analysis easy to use

- Macros are reusable SAS software components that are flexible and modular enough to be applied to a variety use cases. The GITHUB location for the macros in this paper can be found here:

  https://github.com/sasgovernment

  Since macros only require assigning values to input variables in an interface to dynamically change the macros output, they are also easy to use.
- SAS Studio Custom Tasks are point-and-click user interfaces that guide the user through an analytical process. For example, tasks enable users to select as well as fit a spatial regression model. When a user selects a task option, SAS code is generated and runs on the SAS server. Any output (such as graphical results or data) is displayed in SAS Studio.

Integrated and Extensible

Several capabilities in the SAS platform facilitate the integration of the solution as well as make it extensible.

- APIs enable cross-platform integration by allowing applications that are written in various programming languages to communicate by using a standard web-based protocol. This functionality makes it possible for businesses to bridge the gap between different applications and systems.
- Process Flow. Underlying the SAS Platform is the SAS 4GL (Fourth Generation Language) that can be used to extend the capabilities of the SAS Platform to introduce new behaviors or capabilities to fulfill ever-changing requirements. One way to incorporate the SAS 4GL is through a SAS Studio process flow, which consists of one or more objects. Each object is represented by a node in the process flow. The process flow shows the relationship between two or more objects, such as a SAS program, a task, a query, and so on.

Insights

As mentioned, insights can be derived from models and visualizations. Theyare essential toward driving decisions based on data. But what components are needed to generate these insights for the Medicare Opioid use case?

- Clustering. Clustering is a method of data segmentation that puts observations into groups that are suggested by the data. . It does this using an unsupervised machine learning algorithm called K-means clustering. The observations in each cluster tend to be similar in some measurable way, and observations in different clusters tend to be dissimilar. Observations are assigned to exactly one cluster. From the clustering analysis, you can generate a cluster ID variable to use in other models or visualizations. Clusters in the Medicare opioid use case were created to develop three county cohorts – distressed, at-risk, and competitive.
- Spatial Regression. The SPATIALREG (spatial regression) procedure analyzes spatial econometric models for cross-sectional data whose observations are spatially referenced or geo-referenced. For example, Medicare opioid prescribing data that are collected from all 420 Appalachian counties fall into the category of spatially referenced data. Compared to nonspatial regression models, spatial econometric models are capable of handling spatial dependence and spillover effect in a regression setting. The SPATIALREG procedure requires a secondary data set called the Spatial Weights Matrix (WMAT) discussed below in the Methods section. It is useful for identifying spatial clusters or "hotspots" where government agencies can focus their attention.
- County-level maps. SAS Visual Analytics can be extended to include county-level polygons or choropleths for additional insights. County-level polygons allow for insights that would not be discernable with state level analysis. Specifically, they account for significant sub-state variation in data (for example, Virginia) where spatial clusters in one county might differ considerably from clusters elsewhere in the same state. County-level analysis also provides insights on trends and patterns that cut across state boundaries. Appalachia is one very good example: the region is an amalgamation of counties that often have more in common with each other than with other non-

Appalachian counties in the same state.

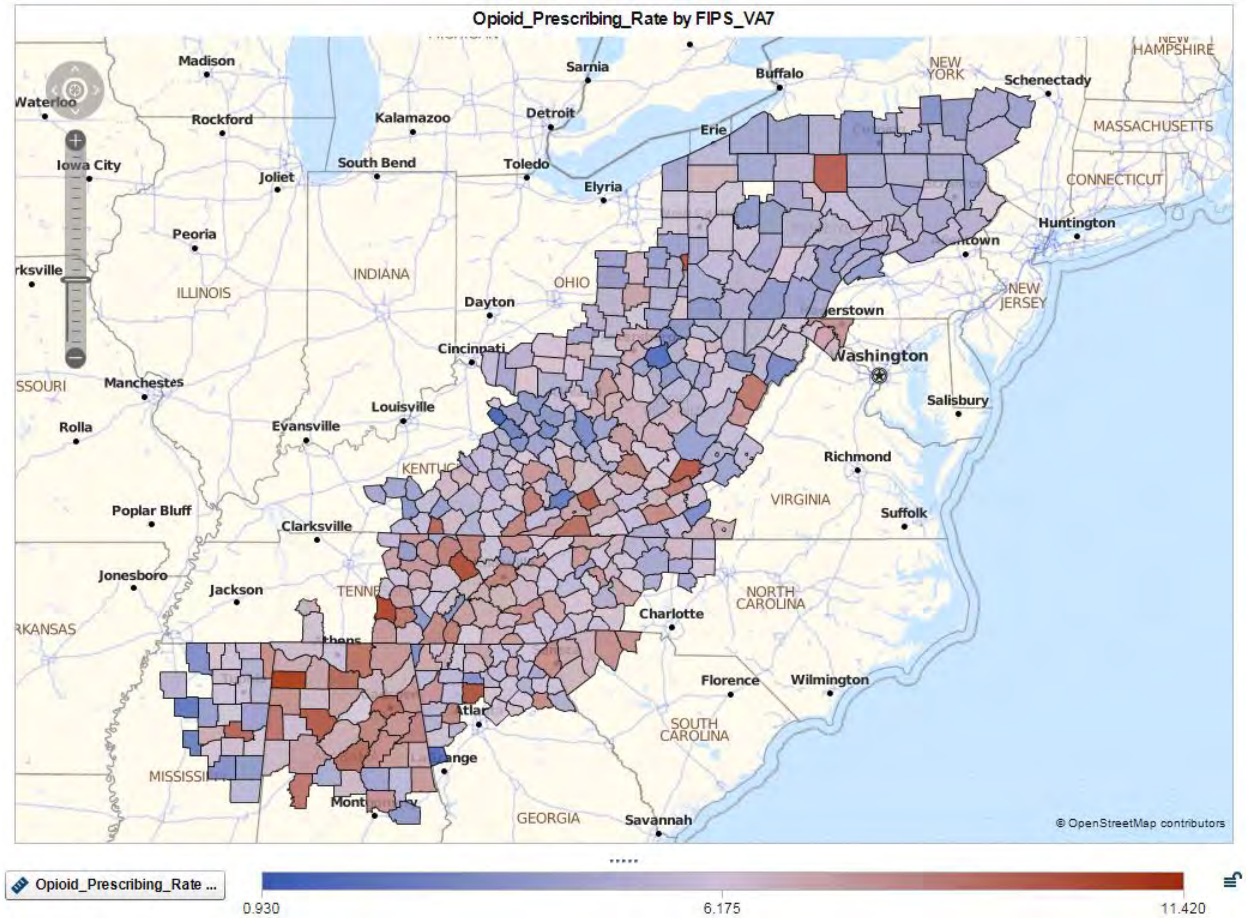

Figure 6. Medicare Part D Opioid Prescribing Rates Shown Using SAS Visual Analytics

Through these capabilities, we will show how to cover all steps of the analytics lifecycle in one integrated environment built on the SAS Platform.

METHOD

We can use this end-to-end solution to answer many analytic questions including the following:

a) What "hotspots" or spatial clusters in Appalachia have a high opioid prescribing rate, which can indicate questionable prescribing practices by Medicare participants?

b) What are the factors that impact opioid prescribing rates?

These questions are based on the following agile user story:

As a (CMS) public health policy maker,

I want to determine which counties are most at risk of improper use of Part D prescriptions or questionable prescribing patterns,

so that I can better target limited (government) investigative resources (federal, state, and local) to those areas to maintain the integrity of the Medicare program and reduce the risk of fraud, waste, or abuse.

The resulting answers and requirements can direct government resources more efficiently to enact change. In addition, understanding the impact of social determinants and economic factors on opioid prescribing rates is crucial for Medicare management and policy making. To answer the two analytic questions above, the research conducted uses a variety of government data from the following sources:

- Chronic Conditions Warehouse (CCW). The CCW is a research database designed to make Medicare, Medicaid, Assessments, and Part D Prescription Drug Event data more readily available to support research designed to improve the quality of care and reduce costs and utilization.
- Behavioral Risk Factor Surveillance System (BRFSS). BRFSS is the nation's premier system of health-related telephone surveys that collect state data about U.S. residents regarding their health-related risk behaviors, chronic health conditions, and use of preventive services. It is run by the Centers for Disease Control and Prevention and conducted by the individual state health departments.
- U.S. Census Bureau. Census helps local officials, government leaders, and businesses understand the changes taking place in their communities. It is the premier source for detailed population and housing information about our nation.
- U.S. Bureau of Labor Statistics (BLS). BLS is the principal fact-finding agency for the federal government in the broad field of labor economics and statistics. The BLS is an independent national statistical agency that collects, processes, analyzes, and disseminates essential statistical data to the American public, the U.S. Congress, other federal agencies, state and local governments, business, and labor
- U.S. Department of Health and Human Services (HHS). HHS data are designed to be used by planners, policy makers, researchers, and others interested in the nation’s health care delivery system and factors that can impact health status and health care in the United States.
- Health Resources and Services Administration (HRSA). The Area Health Resources Files (AHRF) data are designed to be used by planners, policy makers, researchers, and others interested in the nation’s health care delivery system and factors that can impact health status and health care in the United States.

Data collected in this study are geo-referenced since observations are associated with counties and hence are categorized as spatial data. One of the inherent characteristics of spatial data is that observations are often closely correlated with each other and the strength of such correlation depends on the distance between two units in space. Spatial dependence distinguishes the analysis of spatial data from that of nonspatial data. Figure 6 shows Medicare Part D opioid prescribing rates in 2014, which suggests that there is spatial dependence in the data as values in neighboring counties are alike. If we ignore spatial dependence in the data and use regression techniques that are developed for nonspatial data, our parameter estimates and inference will be flawed. As a result, it is very important to have access to dedicated analytic tools when analyzing spatial data. Among the many disciplines that study spatial data, spatial econometrics accounts for spatial dependence and heterogeneity in spatial data in the regression setting [12, 13].

There are three key components in spatial econometric modeling: 1) specification of spatial weights matrix; 2) model specification; 3) and model selection. In spatial econometric modeling, spatial weights matrix plays an important role because it characterizes neighbor relationship between two units and it is used to parameterize different forms of spatial dependence in the data. In practice, spatial weights matrices are created according to various criteria, such as contiguity, distance, and many more.

Model specification involves the choice of explanatory variables that go into the model and the choice of a particular type of model to be fit. Before the modeling process began, the framework shown in Figure 7 was developed to simplify and categorize the variable selection. Possible explanatory variables were segmented by the categories of Medicare, Community and Supply and Demand. The variable selection started like a backward stepwise approach. All variables in each model type were entered into the Spatial

Regression model. Then the variable that failed the t-test by the largest margin was removed. This continued through many iterations until the best explanatory variables rose to the surface.

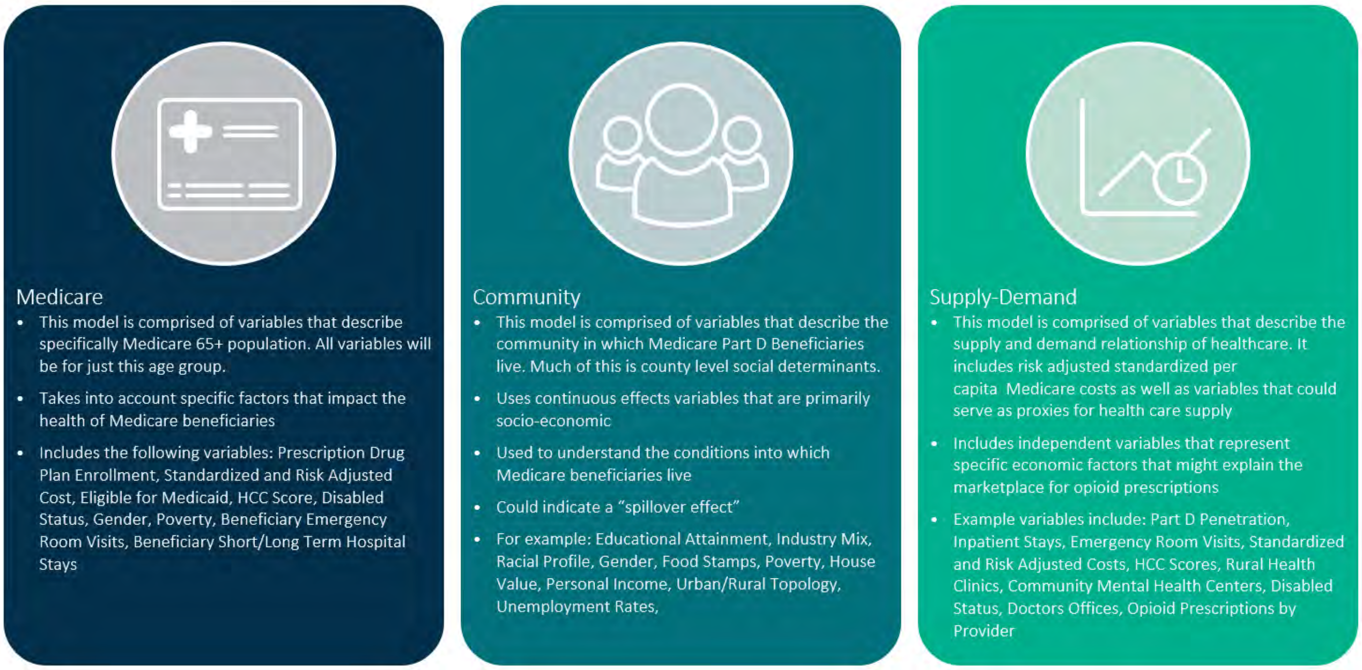

Figure 7. Three Types of Models

To facilitate model specification, it is important to understand three different sources that spatial dependence can arise from: endogenous interaction effect, exogenous interaction effect, and the interaction among the error terms[13]. The endogenous interaction effect means that the value of the dependent variable in one unit is impacted by that of the dependent variables in other units. For exogenous interaction effect, it refers to that the value of the dependent variable in one unit is impacted by that of explanatory variables in other units. In addition, interaction among the error terms describes that the value of the error in one unit is impacted by that of the error in other units.

As is common in many data analyses, we rarely know the true model from which the data at hand is generated. As a result, the spatial econometric analysis flow involves three steps. First, we start with a set of candidate models. Second, we choose the model that best describes the data according to certain criteria, for example, Akaike information criterion (AIC), Bayesian information criterion (BIC), and within-sample mean squared error (MSE). Third, we fit the best model to our data to obtain parameter estimates and draw some conclusions.

To address our analytic questions, we rely on the SPATIALREG procedure for our spatial econometric analysis since spatial econometric models are capable of handling spatial dependence in a regression setting. Dedicated to spatial econometric modeling for Gaussian spatial data [14, 15, 16], the SPATIALREG procedure supports a total of twelve models that are capable of modeling various forms of spatial dependence in the data. To aid with model selection, the SPATIALREG procedure allows users to fit multiple models at once.

In our analysis, the dependent variable is the county-level opioid prescribing rates. To satisfy the normality assumption underlying spatial econometric models, we used the logit transformation for our dependent variable. We consider the AIC criteria for model selection. In other words, the best model is the model with smallest AIC among the models that we considered. In addition, MSE is provided for each model as a way of showing how each model fits the data.

RESULTS

Our goal is to better understand the Medicare beneficiaries, supply and demand (marketplace), and communities surrounding these opioid prescribing rates.

The Medicare/65+ model considers specific factors that impact the health of Medicare beneficiaries. Three factors that we consider are the logarithm of the number of residents enrolled in the Medicare Prescription Drug Plan (*Log_MDCR_PDP_Enroll*), the logarithm of the number of residents with disability(*Log_PctDisable*), and the logarithm of the number of residents who are 65 years old or elder (*Log_Pop65*). The logarithm transformation is considered for these three explanatory variables to ease interpretation and to make measurement comparable.

Figure 8 presents the model selection results for 12 Medicare/65+ models that were considered. According to Figure 8, the spatial Durbin moving average (SDMA) model has the smallest AIC and hence is identified to be the winning model. In addition, the SDMA model also fits the data well relative to other models according to MSE.

Obs	Model	AIC	MSE
1	SDMA	778.94229	0.36514
2	SDM	779.10150	0.36508
3	SDEM	779.22902	0.36514
4	SDARMA	780.93181	0.36520
5	SDAC	781.06728	0.36508
6	SMA	784.32313	0.37708
7	SARMA	784.45320	0.37704
8	SEM	784.63757	0.37711
9	SLX	784.72875	0.36510
10	SAR	785.74111	0.37740
11	SAC	786.39573	0.37703
12	Linear	791.45114	0.37633

Figure 8. Model Selection Results for Medicare

The parameter estimates from the SDMA model are shown in Figure 9. According to Figure 9, the significant findings from this model are:

- Each of the two variables Log_PctDisable and Log_Pop65 has a significant positive indirect effect on opioid prescribing rates at the 5% level.
- The variable Log_MDCR_PDP_Enrll has a significant positive direct effect and a significant negative indirect effect on opioid prescribing rates, both at the 5% level.
- The parameter _lambda is significant negative at the 5% level, which indicates that there is positive spatial dependence in the error terms after accounting for three explanatory variables.

Model: MODEL 2
Dependent Variable: Logit_OpioidRate

Model Fit Summary	
Dependent Variable	Logit_OpioidRate
Number of Observations	420
Data Set	WORK.SGF_OPIOID_APPLACHIAN0_CL_AHRF
Spatial Weights	WORK.APPLACHIAN_WLIST
Model	SDMA
Log Likelihood	-380.47114
Maximum Absolute Gradient	1.52785E-6
Number of Iterations	16
Optimization Method	Newton-Raphson
AIC	778.94229
SBC	815.30458

Algorithm converged.

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	-8.164892	1.704265	-4.79	<.0001
Log_MDCR_PDP_Enrll	1	0.528968	0.195028	2.71	0.0067
Log_PctDisable	1	-0.325727	0.272294	-1.20	0.2316
Log_Pop65	1	-0.376959	0.198627	-1.90	0.0577
W_Log_MDCR_PDP_Enrll	1	-0.637849	0.284230	-2.24	0.0248
W_Log_PctDisable	1	1.304656	0.417005	3.13	0.0018
W_Log_Pop65	1	0.672368	0.284314	2.36	0.0180
_lambda	1	-0.230960	0.086909	-2.66	0.0079
_sigma2	1	0.362021	0.025128	14.41	<.0001

Figure 9. Parameter Estimates from the SDMA Model for Medicare

The Suppy and Demand model considers specific economic factors that might explain the marketplace for opioid prescriptions. The economic factors that we include in this model are: the logarithm of Medicare Fee for Service beneficiary emergency room visits in 2014 (Log_MDCR_ER_VISIT), the logarithm of the average HCC score in 2014 (Log_Avg_HCC), and the logarithm of the number of residents with disability (Log_PctDisable).

Figure 10 presents the model selection results for twelve Supply and Demand models that we considered. According to Figure 10, the spatial Durbin autoregressive moving average (SDARMA) model has the smallest AIC. However, this model has convergence issues and hence is excluded from model selection. As a result, we choose the spatial Durbin model (SDM) to be the best model since it has the smallest AIC among the remaining eleven models. In addition, the SDM also fits the data well relative to other models according to MSE.

Obs	Model	AIC	MSE
1	SDARMA	743.26649	0.35199
2	SDM	779.45604	0.36289
3	SDMA	779.78921	0.36370
4	SDEM	780.18185	0.36369
5	SDAC	781.36335	0.36253
6	SLX	783.04381	0.36364
7	SMA	784.60568	0.37535
8	SEM	784.88905	0.37542
9	SAR	784.88939	0.37468
10	SARMA	786.02999	0.37848
11	SAC	786.82358	0.37498
12	Linear	789.66659	0.37474

Figure 10. Model Selection Results for Supply and Demand Models

The parameter estimates from the SDM are shown in Figure 11. According to Figure 11, the significant findings from this model are:

- The log of Medicare Beneficiary Emergency Room Visits was found to have a significant positive short run direct effect on opioid prescribing rates at the 5% level.
- The parameter _rho is significant positive at the 5% level, which indicates that there is a positive spatial dependency in the data.

Model: MODEL 2
Dependent Variable: Logit_OpioidRate

Model Fit Summary	
Dependent Variable	Logit_OpioidRate
Number of Observations	420
Data Set	WORK.SGF_OPIOID_APPLACHIAN0_CL_AHRF
Spatial Weights	WORK.APPLACHIAN_WLIST
Model	SDM
Log Likelihood	-380.72802
Maximum Absolute Gradient	4.36977E-6
Number of Iterations	13
Optimization Method	Newton-Raphson
AIC	779.45604
SBC	815.81834

Algorithm converged.

Parameter	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	-6.326160	1.321043	-4.79	<.0001
Log_MDCR_ER_Visit	1	0.156549	0.040157	3.90	<.0001
Log_MDCR_AVG_HCC	1	0.249479	0.656387	0.38	0.7039
Log_PctDisable	1	-0.422021	0.278836	-1.51	0.1302
W_Log_MDCR_ER_Visit	1	0.032637	0.068311	0.48	0.6328
W_Log_MDCR_AVG_HCC	1	-1.874150	0.981277	-1.91	0.0561
W_Log_PctDisable	1	1.064079	0.359883	2.96	0.0031
_rho	1	0.179481	0.074524	2.41	0.0160
_sigma2	1	0.356486	0.024681	14.44	<.0001

Figure 11. Parameter Estimates from the SAS Deployment Manager Model for Supply and Demand

The Community model uses social determinant factors to explain opioid prescribing rates. We consider four social determinant factors: the logarithm of the percentage of people with a Bachelor's degree or above (*Log_Bachelors*), the logarithm of the percentage of people living in poverty (*Log_Poverty_Per*), the logarithm of the percentage of people used in Education, Health or Social Services (*Log_Pct_EduHlth*), and the logarithm of the percentage of people used in Manufacturing (*Log_Pct_Manu*).

Figure 12 presents the model selection results for twelve community models that we considered. According to Figure 12, the SDM has the smallest AIC. Based on the MSE, we note that the SDM model fits the data well due to its small MSE relative to other models.

Obs	Model	AIC	MSE
1	SDM	768.68818	0.34927
2	SDARMA	769.25926	0.34680
3	SDMA	769.55626	0.35037
4	SDEM	769.87056	0.35036
5	SDAC	770.45211	0.34873
6	SLX	771.33859	0.35030
7	SARMA	785.01446	0.37732
8	SMA	789.10966	0.37905
9	SEM	789.39558	0.37925
10	SAC	789.95917	0.38078
11	SAR	790.23984	0.37795
12	Linear	794.78456	0.37753

Figure 12. Model Selection Results for Community Models

The parameter estimates from the SDM model are shown in Figure 13. According to Figure 13, the significant findings from this model are:

- Both variables *Log_Bachelors and Log_Pct_Manu have a significant positive short run direct impacts on* opioid prescribing rates at the 5% level.
- The parameter _rho is significant and positive at the 5% level, which indicates that there is a positive spatial dependency in the data.

Model: MODEL 2
Dependent Variable: Logit_OpioidRate

Model Fit Summary	
Dependent Variable	Logit_OpioidRate
Number of Observations	420
Data Set	WORK.SGF_OPIOID_APPLACHIAN0_CL_AHRF
Spatial Weights	WORK.APPLACHIAN_WLIST
Model	SDM
Log Likelihood	-373.34409
Maximum Absolute Gradient	4.53436E-6
Number of Iterations	13
Optimization Method	Newton-Raphson
AIC	768.68818
SBC	813.13098

Algorithm converged.

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	-2.199972	1.216011	-1.81	0.0704
Log_Bachelors	1	0.152730	0.032211	4.74	<.0001
Log_Poverty_Per	1	-0.142063	0.160818	-0.88	0.3770
Log_Pct_EduHlth	1	0.133261	0.209838	0.64	0.5254
Log_Pct_Manu	1	0.318175	0.092987	3.42	0.0006
W_Log_Bachelors	1	-0.020737	0.053033	-0.39	0.6958
W_Log_Poverty_Per	1	0.762899	0.230684	3.31	0.0009
W_Log_Pct_EduHlth	1	-0.958702	0.324200	-2.96	0.0031
W_Log_Pct_Manu	1	-0.493774	0.129719	-3.81	0.0001
_rho	1	0.164986	0.075258	2.19	0.0284
_sigma2	1	0.344528	0.023841	14.45	<.0001

Figure 13. Parameter Estimates from the SDM for Community

To assess the in-sample predictive performance of our models, we compute on the ratio of predicted opioid prescribing rates to observed opioid rates. For the three best models identified by AIC, the plots for the ratio of predicted opioid prescribing rates to observed opioid rates have identical color scheme for the ratio scale provided in Figure 15. According to Figure 14, we conclude that all three models fit the data fairly well. Furthermore, we note some poor predictions associated with counties on the borders, which might be due to the edge effect.

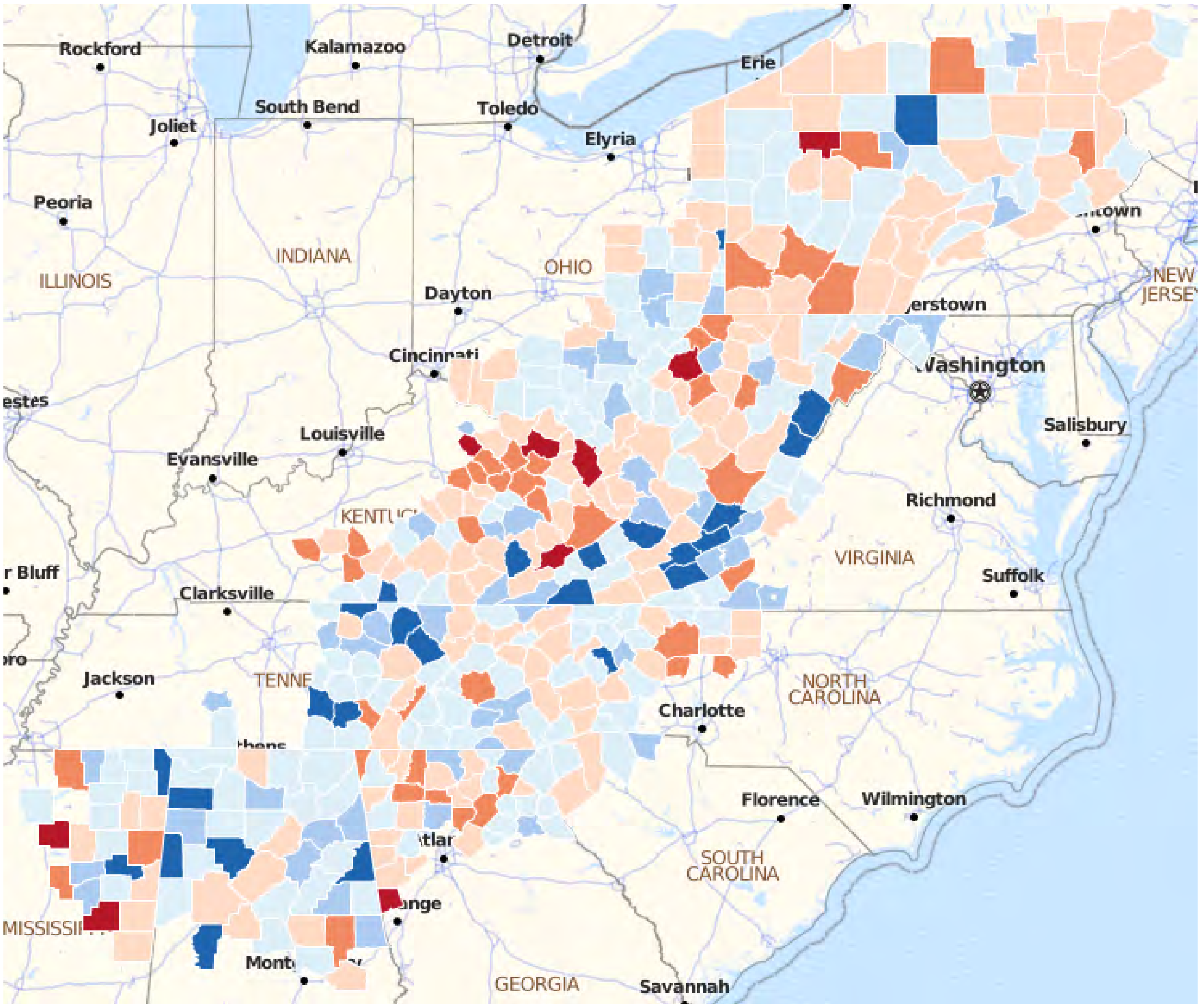

Figure 14. Plot of the Ratio of Predicted to Observed Opioid Prescribing Rates

Color	Color Name	Value Range	Prediction Quality
	Red	Ratio<.8	Poor
	Orange	.8-.9	Fair
	Light Orange	.9-1	Good
	Light Blue	1-1.1	Good
	Blue	1.1-1.2	Fair
	Dark Blue	Ratio>1.2	Poor

Figure 15. Color Scheme for the Ratio of Predicted to Observed Opioid Prescribing Rates

CONCLUSIONS

Although the ‘true’ model for opioid prescribing rates will always remain elusive, these three models provide us with different lenses to look at the opioid prescription rate in the Medicare Part D population. The results suggest that for the Medicare, Supply and Demand, and Community models that we considered, spatial econometric models outperform the purely linear regression models according to the model selection criterion.

From the Medicare/65+ model, we found that both Log_Pop65 and Log_PctDisable have a significant indirect effect on opioid prescribing rates at 5% level, which indicates local spillover effects. In other words, the opioid prescription rate in one county is affected by the number of residents who are 65 years old or elder and the number of residents with disability in its neighboring counties. Moreover, the number of residents enrolled in the Medicare Prescription Drug Plan is found to have a positive direct effect on opioid prescribing rates.

From the Supply and Demand model, we understand that Medicare emergency room visits has a short run direct impact on the opioid prescription rate. In addition, the best model identified according to AIC suggest positive spatial dependence in the data, which necessitates the need for spatial econometric modeling.

The Community model showcases the need for spatial regression in our analysis due to positive spatial dependence in the data. We also found that counties with more educated residents and more residents employed in manufacturing tend to have higher opioid prescribing rates.

The Supply and Demand and Community models deserve particular attention since they suggest that economic and social determinants of health play a particularly significant role in opioid prescribing rates (as noted by their AIC result). The indirect effect of the percentage of people living below the poverty line on opioid prescribing rates would have remained hidden if not for spatial regression techniques. Further research is required to determine whether the low-income subsidy benefit is working as desired, however the results shown above provide some evidence that it is impacting high poverty areas.

Armed with the analytics lifecycle, spatial regression techniques and these insights, Medicare policy makers have a framework to analyze and understand opioid prescribing rates within a geospatial context to target limited government resources efficiently.

BIBLIOGRAPHY

1. HHS OIG Data Brief • July 2017 • OEI-02-17-00250. Opioids in Medicare Part D: Concerns about Extreme Use and Questionable Prescribing

   https://oig.hhs.gov/oei/reports/oei-02-17-00250.pdf

2. https://www.newyorker.com/magazine/2017/10/30/the-family-that-built-an-empire-of-pain
3. https://www.wvgazettemail.com/news/health/opioid-epidemic-costs-wv-billion-annually-study-says/article_1cd8aaa5-78eb-5fd5-8619-3a0a1c086e66.html
4. https://www.kff.org/disparities-policy/issue-brief/beyond-health-care-the-role-of-social-determinants-in-promoting-health-and-health-equity/
5. Figallo-Monge, Manuel. 2016. “Pedal-to-the-Metal Analytics with SAS® Studio, SAS® Visual Analytics, SAS® Visual Statistics, and SAS® Contextual Analysis.” Proceedings of the SAS Global Forum 2016 Conference. Cary NC: SAS Institute Inc. Available http://support.sas.com/resources/papers/proceedings16/SAS6560-2016.pdf
6. Baer, Darius Ph.D., and Suneel Grover, MBA, MS. 2016. Enhanced Segmentation Using SAS® Visual Analytics and SAS® Visual Statistics. SAS Institute Inc. Available: http://support.sas.com/resources/papers/proceedings16/SAS6222-2016.pdf
7. Chen, Li-Hui PhD and Manuel Figallo. Bridging the Gap: Importing Health Indicators Warehouse Data into SAS® Visual Analytics Using SAS® Stored Processes and APIs. Consumer Product

Safety Commission; The SAS Institute Inc. Available: http://support.sas.com/resources/papers/proceedings16/10540-2016.pdf

8. Chen, Li-Hui, Manuel Figallo, and Josh McCullough. Bridging the Gap: Importing Health Indicators Warehouse Data into SAS® Visual Analytics using Stored Processes and APIs. National Center for Health Statistics, Centers for Disease Control and Prevention; SAS Institute Inc.; Social & Scientific Systems Inc. Available:

   http://analytics.ncsu.edu/sesug/2015/RV-92.pdf

9. Figallo-Monge, Manuel. Macro Design and Usage in a Multi-Tier Architecture for ETL and Google Visualization API Integration, DevTech Systems, Inc. Available: http://support.sas.com/resources/papers/proceedings12/003-2012.pdf

10. Figallo-Monge, Manuel. Using Design Patterns to Integrate SAS/IntrNet® with Web Technologies. Freddie Mac. Available:

    http://www2.sas.com/proceedings/sugi31/007-31.pdf

11. Sabo, Tom. 2017. "Applying Text Analytics and Machine Learning to Assess Consumer Financial Complaints." Proceedings of the SAS Global Forum 2017 Conference. Cary NC: SAS Institute Inc. Available http://support.sas.com/resources/papers/proceedings17/SAS0282-2017.pdf

12. Anselin, L. (1988). Spatial Econometrics: Methods and Models. Amsterdam: Springer.

13. Elhorst, J. P. (2013). Spatial Econometrics: From Cross-Sectional Data to Spatial Panels. Berlin: Springer.

14. Wu, G., and Chvosta, J. 2016. "How Do My Neighbors Affect Me? SAS/ETS Methods for Spatial Econometric Modeling." Proceedings of the SAS Global Forum 2016 Conference. Cary, NC: SAS Institute Inc.

15. Wu, G., and Chvosta, J 2017. "Big Value from Big Data: SAS/ETS Methods for Spatial Econometric Modeling in the Era of Big Data." Proceedings of the SAS Global Forum 2017 Conference. Cary, NC: SAS Institute Inc.

16. SAS Institute Inc. 2016. SAS/ETS® 14*.2 User's Guide*, Cary, NC: SAS Institute Inc.

17. https://datawarehouse.hrsa.gov/topics/ahrf.aspx

18. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4417427/

19. https://www.arc.gov/appalachian_region/TheAppalachianRegion.asp

ACKNOWLEDGMENTS

The authors would like to express their gratitude to their friends and family for supporting them during the writing of this paper. In particular, our uniquely congenial colleagues at SAS: Emily McRae, Matt McTiernan, Sree Vuthaluru, and Kelsey Young who played critical roles in the original build-out for the Health Care Analytics Solution on which this paper is based. Adam Maness, Brian Nichols, Steve Kearney, Evans Travis, Paul Billy for their continued support and guidance. And, without our customers at The U.S. Department of Health and Human Services (HHS), this paper simply would not have been possible: Lily Chen, Frederick Deese, Ray Mierwald, Shanika Palm, Jonathan Smith, and George Wilson.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Manuel Figallo
SAS Institute, Inc. 1530 Wilson Blvd, Suite 800 Arlington, VA 22201
Manuel.Figallo@sas.com
https://www.linkedin.com/in/mfigallo

Guohui Wu
SAS Institute, Inc. 100 SAS Campus Drive, Cary NC 27513
Guohui.Wu@sas.com

Jonathan McCopp
SAS Institute, Inc. 1530 Wilson Blvd, Suite 800 Arlington, VA 22201
Jonathan.McCopp@sas.com
https://www.linkedin.com/in/jonathanmccopp/

APPENDIX A

The health economics research in this paper is part of a larger Health Care Analytics for Government framework.