Clustering Analysis of Statewide Health Insurance Marketplaces

Wanlin Chen, Ian Davis, Lu Shen, Micky Sun,

The Department of Economics, University of Southern California

Econ 570: Big Data Econometrics

Professor Timothy Armstrong

December 13, 2022

Introduction

In 2014, the Affordable Care Act (ACA), also known as Obamacare, was passed to regulate the health insurance markets by providing a marketplace for individuals to purchase health insurance plans. President Obama stated that there should be “a market where Americans can one-stop shop for healthcare plan”. None of these plans can deny coverage on the basis of a preexisting condition and all these plans include an affordable basic benefit package that includes prevention, and protection against catastrophic costs (2009). In accordance with the legislation, each state has developed a health insurance market allowing its citizens to purchase their own plans. All insurance plans on the marketplace are regulated by the ACA and have mandatory attributes such as not requiring increased cost sharing for pre-existing conditions and covering primary immunizations. Plans in marketplaces are categorized as four metal levels: bronze, silver, gold and platinum. The metal level helps identify different premiums and the deductible amounts. Other than metal levels, marketplace health insurance plans also have various attributes like copayment and coinsurance. With the marketplaces, large numbers of individuals are able to shop for health insurance with explicit policies based on their need for healthcare services. The Congressional Budget Office and Joint Committee on Taxation estimate, in 2022, 1.7 million people would gain coverage through the ACA marketplaces, 1.3 million of whom were previously uninsured.

This research explores a 2019 plan-level health insurance dataset from the United States to analyze geographic variations in health insurance marketplaces. We hypothesized that insurance plans exhibit heterogeneous group attributes across the states resulting from different characteristics among states including demographics, provider practices, and ideologies. The study leverages a state-level clustering analysis to find how health insurance markets relate to each other and mirrors the results found in the literature. The research shows that there exist large differences in plan policies in the marketplaces between states, aligning with the initial hypothesis. The research also makes attempts to predict state health insurance coverage rate on a contemporary level. The result implies that although large variations exist in health insurance structure, they do not have a large effect on evidence based care.

Literature review

Previous studies put forward that medical resources are distributed unevenly creating quality and capacity care gaps across the states. Miami is especially sufficient in effective care capacity, having 229.7 physicians per 100,000 people while Salt Lake City has only 155.6 physicians per 100,000 people in 1996. Patients in the last six months of life visit the ICU 49.3% of the time in Miami, recorded in 1995. Similar enrollees in Salt Lake City are admitted only 20.6% of the time, less than half of the rate in Miami (Fisher, 2003). Song, etc (2010) conducted a study using Medicare claims data from 1999 through 2006, finding that the medical resources utilized increased among beneficiaries who moved from regions with a higher intensity of practice than among those who moved to regions with the same or lower intensity of practice. Hence this study investigates the marketplaces health insurance attributes and tries to expand the literature by showing these marketplaces exhibit similar differences in quality and quantity of care covered between states.

A rich body of literature has emerged analyzing ACA marketplaces which has clearly improved health care outcomes, especially among low-income adults (Goldman et al., 2018). Notably, previous research revealed that healthcare insurance enrollment varied across rural areas. Of the explanatory variables examined in the model, state-level policies, specifically

Medicaid expansion, were most associated with enrollment (Drake et al., 2015). The study of changing healthcare insurance market dynamics indicated that plan shifts can negatively affect the future. Future healthcare policies should be devoted to stabilizing the insurance market (McKillop et al., 2018). Chen and Page (2018) provided a comprehensive insight into how insurance plan deductible levels affect the healthcare experience, showing that healthcare accessibility, affordability, routine checkups, out-of-pocket costs, and satisfaction were adversely affected by the deductible.

Data

Information about insurance plans offered in state marketplaces come from the 2019 Plan Attribute Public Use File provided by the Centers for Medicare and Medicaid Services' Center for Consumer Information and Insurance Oversight (CCIIO). The dataset is described by CCIIO as a "Plan-level dataset on maximum out of pocket payments, deductibles, HSA eligibility, and other plan attributes." One row of the dataset represents an individual insurance plan offered in a state-based health insurance exchange (CMS (n.d)). In order to adjust the dataset from policy level to state level, most variables had to be reformatted to represent the average of the variable within a given state. Indicator variables were transformed into probabilities of a given health plan within a state having an associated attribute. Finally, plans only offering dental services were dropped and variables with state level missing values were dropped as well. The former allows for a more precise comparison between states and the latter is a requirement for the clustering algorithm employed.

Methods

The goal of the analysis is to identify underlying state groupings based off of the attributes from the insurance policies offered in their exchanges. To accomplish this objective, we employ a K-means clustering algorithm which allows for us to partition our data into K unique clusters with the objective of minimizing the within-cluster variation (James, 2022).

The algorithm assumes a given data set $X_i \; \forall \; i \in \{1, N\}$ where $X_i = \left[X_{i1}, \ldots, X_{id}\right]'$ has K underlying, non-overlapping, groups. To start, each X is randomly assigned a cluster $C_j \ni j \in \{1, K\}$ and cluster centers are then estimated. The cluster center can be thought of as the average X within the cluster

$$\widehat{\mu}_k = \frac{\sum_{i=1}^{n} X_i 1\{\widehat{C}_i = k\}}{\sum_{i=1}^{n} 1\{\widehat{C}_i = k\}}$$

Each observation is then placed in the cluster where $\sum_{m=1}^{d} (X_{ij} - \widehat{\mu}_{kj})^2$ is the smallest. In words, the observation is placed in the cluster in which the cluster average has the smallest euclidean distance. Cluster centers are then recalculated and observations are resorted until a solution is approximated.

Results

Data Cleaning Results

The original dataset contained 15,690 insurance plans offered across 39 states. 2,077 plans only covering dental services were dropped leaving 13,613 plans in the dataset. Variables were reformatted to read as averages and proportions within each state. The variable names, their descriptions, means, and ranges are shown in **Table 1** below.

Variable	Description	Mean	Range
st_policycount	Count of policies offered within a state	193	[29,1484]
state_ppo	Probability a plan offered in a state is a preferred provider organization plan	.42	[0,1]
state_pos	Probability a plan offered in a state is a point of service plan	.06	[0,1]
state_epo	Probability a plan offered in a state is an exclusive provider organization	.27	[0,1]
state_platinum	Probability a plan offered in a state is a platinum rated plan	.02	[0,.14]
state_gold	Probability a plan offered in a state is a gold rated plan	.16	[.08,.24]
state_silver	Probability a plan offered in a state is a silver rated plan	.52	[.21,.8]
state_bronze	Probability a plan offered in a state is a bronze rated plan	.18	[0,.37]
state_catastrophic	Probability a plan offered in a state offers catastrophic coverage	.3	[0,.08]
state_pregnancy_noti-ce	Probability a plan offered in a state requires notice for pregnancy coverage	.2	[0.1]
state_prior_auth_spec	Probability a plan offered in a state requires prior authorization for care from specialists	.16	[0,.72]
state_wellness	Probability a plan offered in a state offers a wellness program	.33	[0,1]
state_asthma	Probability a plan offered in a state offers an asthma management program	.9	[0,1]
state_heardisease	Probability a plan offered in a state offers a heart disease management program	.84	[0,1]
state_depression	Probability a plan offered in a state offers a depression management program	.55	[0,1]
state_diabetes	Probability a plan offered in	..93	[0,1]

	a state offers a diabetes management program		
state_hbd_hc	Probability a plan offered in a state offers a high blood pressure and high cholesterol management program	.77	[0,1]
state_lowerbackpain	Probability a plan offered in a state offers a low back pain management program	.54	[0,1]
state_painmanagement	Probability a plan offered in a state offers a pain management program	.47	[0,1]
state_pregnancy	Probability a plan offered in a state offers a pregnancy management program	.63	[0,1]
state_weightlossprogr-ams	Probability a plan offered in a state offers a weight loss program	.59	[0,1]
state_outofservice	Probability a plan offered in a state provides out of service area coverage	.68	[0,1]
state_nationalnetwork	Probability a plan offered in a state is a part of a national network	.3	[0,1]
state_av	Average actual value of the plan issuer within a state	.78	[.72,.83]
state_mult_net_tiers	Probability a plan offered in a state has two network tiers	.18	[0,.74]
state_first tierutil	Average projected first tier utilization within a state	94.34	[70.12,100]
state_secondtierutil	Average projected second tier utilization within a state	5.6	[0.29.88]
state_babydeduct	Average projected deductible for having a baby within a state	200.8	[.1.98,533.79]
state_babycoin	Average projected coinsurance for having a baby within a state	1249.13	[584.71,2361.29]
state_babylim	Average projected limit for a having a baby within a state	62.88	[0,242.4]
state_diabetesdeduct	Average projected deductible for diabetes care within a state	1661.1	[783.77,2806.5]
state_diabetescopay	Average projected copay for	683.9	[122.22,1341.44]

	diabetes care within a state		
state_diabetescoin	Average projected coinsurance for diabetes care within a state	402.72	[60.29,980.13]
state_diabeteslim	Average projected limit for diabetes care within a state	59.52	[0,275.588]
state_fracdeduct	Average projected deductible for a simple bone fracture within a state	909.97	[547.16,1424]
state_fraccopay	Average projected copay for a simple bone fracture within a state	177.94	[23.44,1044.39]
state_fraccoin	Average projected coinsurance for a simple bone fracture within a state	128.53	[24.5,322.89]
state_fraclim	Average projected limit for a simple bone fracture within a state	1.59	[0,62.02]
state_begprimcostsha-re	Average amount of primary care visits fully covered before costsharing begins within a state	.098	[0,.62]
state_begprimededuct	Average amount of primary care visits with copay allowed before subject to deductible/out of pocket limits	.25	[0,2.77]
sttinntier1individual-moop	Average amount of the tier 1 in network, individual out-of-pocket cost limit for medical and drug essential health benefits	5018.41	[4351.91,6513]
sttinntier1familyperpe-rsonmoop	Average amount of the tier 1 in network, family per person out-of-pocket cost limit for medical and drug essential health benefits	5031.12	[4351.91,6513]
sttinntier1familypergroupmoop	Average amount of the tier 1 in network, family per group out-of-pocket cost limit for medical and drug essential health benefits	100,38	[8,703.82,130,26]
sttdedinntier1individu-al	Average individual deductible for tier 1 in network medical and drug essential health benefits	3229.15	[1886.43,5053.1]
sttdedinntier1familyp-erpers	Average family per person	3218.96	[1886.43,5053.1]

on	deductible for tier 1 in network medical and drug essential health benefits		
sttdedinntier1familyp-ergroup	Average family per group deductible for tier 1 in network medical and drug essential health benefits	6635.7	[3419.12,10106.21]
state_hsa	Probability a plan qualifies for a health savings account within a state	.10	[0,.28]

Table 1: Variable Names, Descriptions, Means, and Ranges

To illustrate the variations of plan attributes found in state marketplaces, average projected copay for childbirth and average projected copay for diabetes related care are shown in **Figure 1** and **Figure 2**.

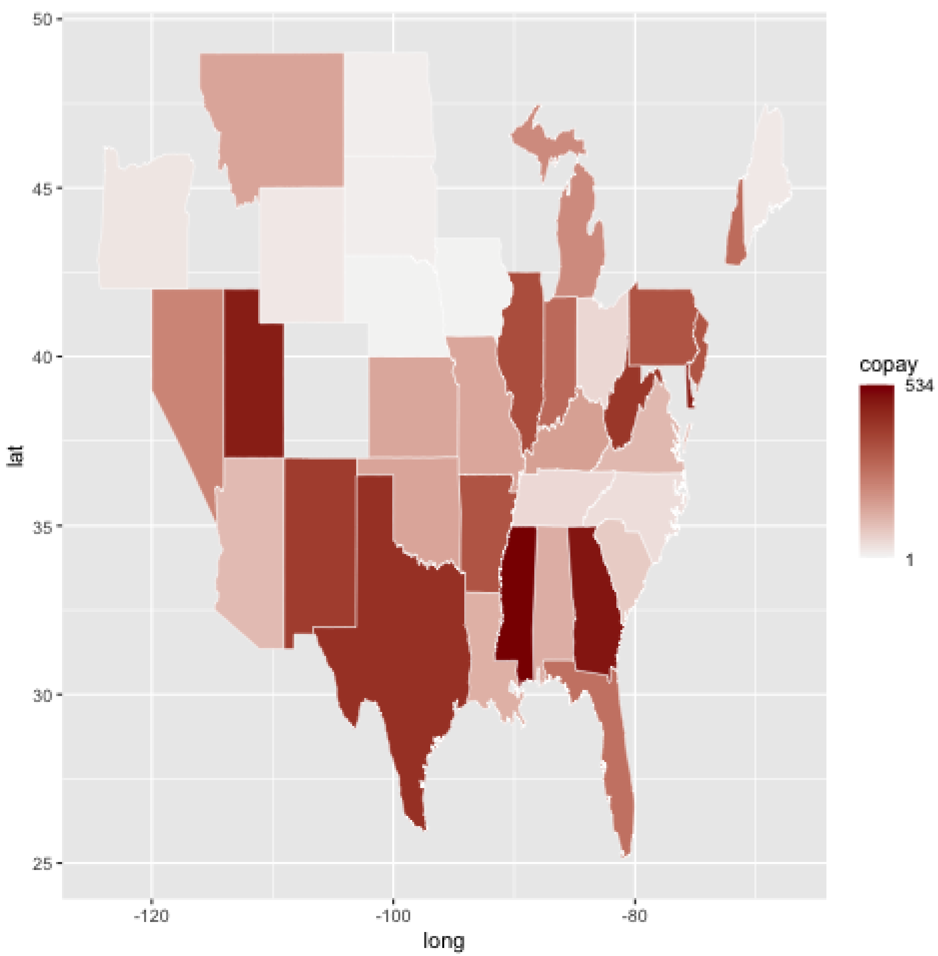

Figure 1: Average Projected Copay for Having a Baby

There is a great deal of variation in projected copay for having a baby with some regional patterns forming. The most striking of which comes from the states Iowa, Nebraska, South Dakota, North Dakota, and Wyoming all making up the northern center of the country and all of low projected copays. Geographic areas that do not lend themselves to patterns remain of interest as well. Alabama has a strikingly low projected deductible compared to its neighbors, Mississippi and Georgia. This low projection is beaten out further by the rest of the southeast states including Tennessee, North Carolina, and South Carolina.

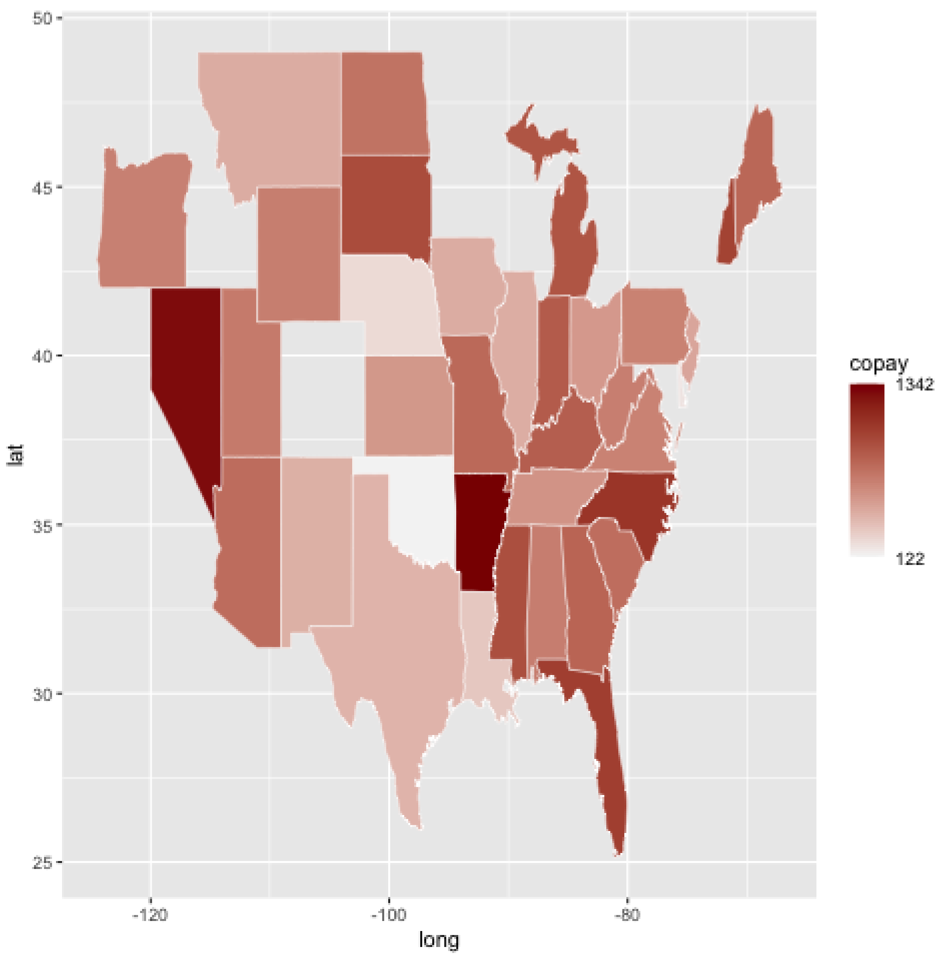

Figure 2: Average Projected Copay for Diabetes Related Care

The average projected copay for diabetes related care tells a different but equally interesting story to that of projected copays. There is a much larger range of values across states with the highest copay states dwarfing their neighbors. Most notably is Arkansas with an average copay of $1,341.44 neighboring states such as Louisiana (average copay of $310.10) and Oklahoma (average copay of $122.22)

Algorithm Results

Four clusters were used for this analysis and resulted in within cluster sum of squares of 480.62, 57.6, 567.84, and 267.35. The between sum of squares divided by the total sum of squares is equal to .277 which indicates that the assigning of four clusters to the dataset as opposed to 39 (original sample size) reduced the sum of squares by 27.7%. The clusters are visualized **Figure 3** below.

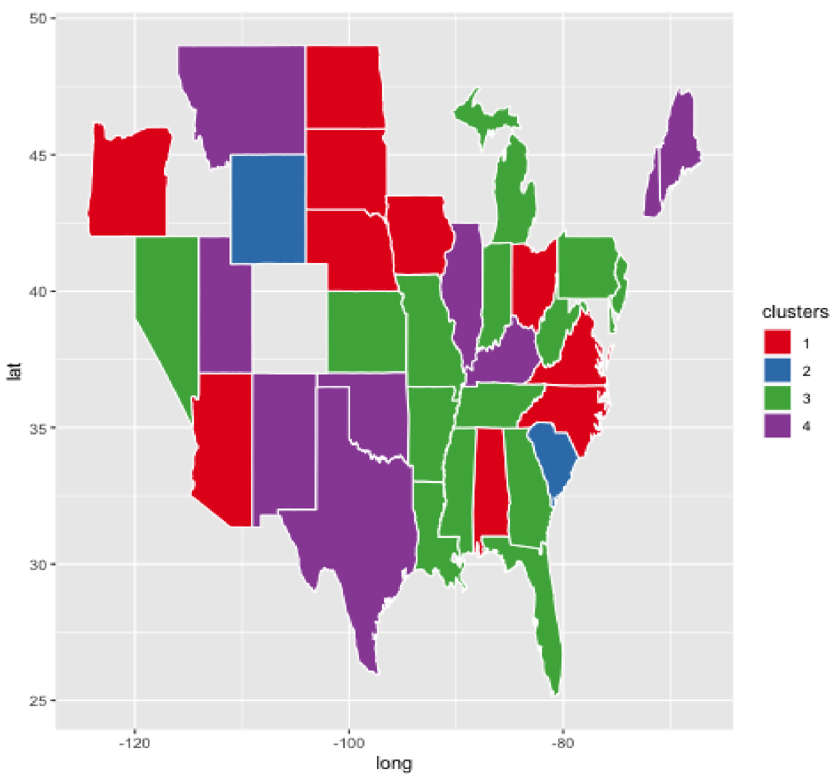

Figure 3: State Clusters

Additional Analysis

From the results of K-means clustering, we showed insurance plan attribute variation exists across states. Despite the clusters shown in the US map revealing the geographic pattern to some extent, there still is the fact that huge gaps appear between neighboring states which break the pattern in those regions. In addition to geographic patterns, something in the ideological nature among different states could contribute to forming another pattern of insurance plan attribute variation. Therefore, we investigate further on the impact of ideology of states by tweaking the parameters of clustering, grouping the states into two clusters, the liberal and the conservative. The political ideology data comes from the 2019 State & Legisative Partisan Composition Table provided by the National Conference of State Legislatures (NCSL).

In **Figure 4**, the cluster assignment is plotted by picking the probability a state will offer a pregnancy management program on the x axis with the degree of liberalism in a state on the y axis and coloring observations in red for the conservative and in blue for the liberal. There are obvious gatherings of the two clusters that, except for the slightly blurred boundary between the liberal and the conservative states in the longitudinal direction, the probabilities that a state will offer a pregnancy program of the liberal states colored in blue are all above 0.4 gathering in the upper-right-hand part of the quadrant. Additionally, the distribution of the conservative states colored in red is evenly located in the horizontal direction. This result does make sense that a state in the red cluster will be more likely to support the Republican Party which is less likely to prioritize creating an expansive health insurance exchange. On the contrary, a state in the blue cluster will be more likely to support the Democratic Party so that the insurance marketplace has more funding which increases the scale of the markets. This increase in scale also increases the number of competing firms/plans which could cause all insurers to offer more generous plans.

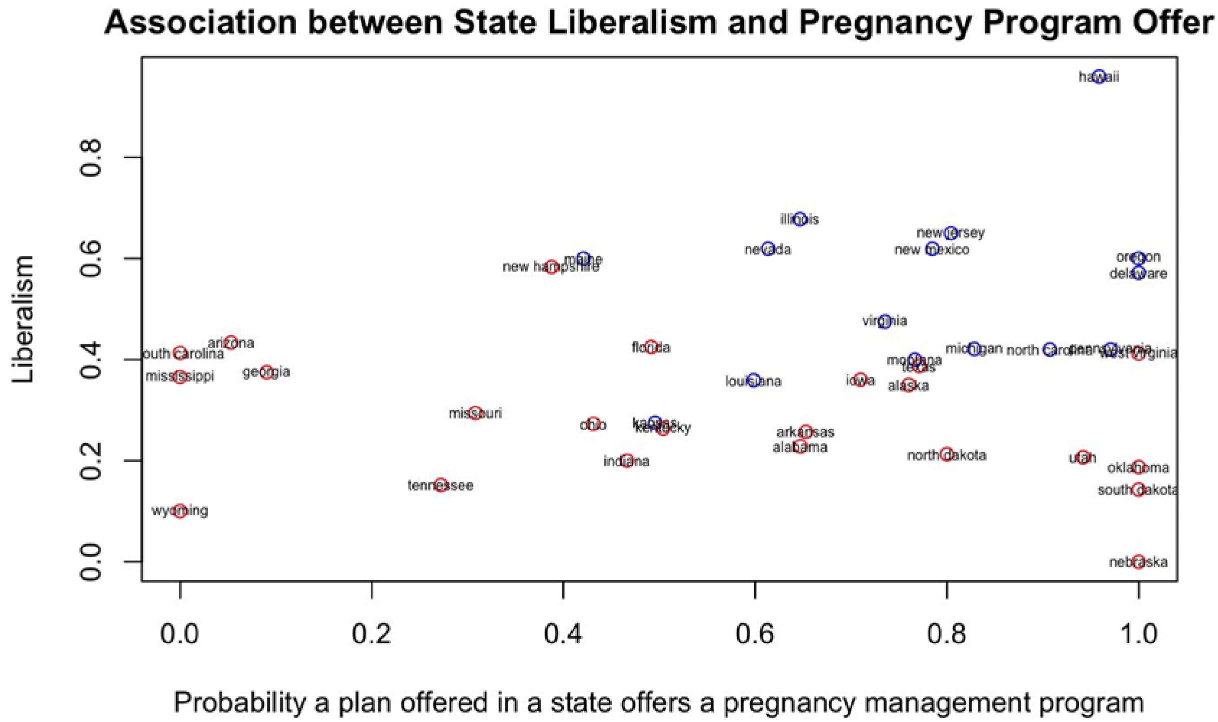

Figure 4 Probability of Plans Offering Pregnancy Management Program

Conclusion

From the results of K-means clustering above, we confirm our hypothesis that insurance plans exhibit heterogeneous group attributes across the states resulting from different characteristics among states including demographics and ideologies. In the first clustering analysis, the states with similar health insurance policies, the results indicate that there exist large disparities and geographic patterns of insurance plan policies on the marketplaces among states, consistent with existing literature. In the extra analysis for the states with similar ideology, the result reflects the impact of the political nature of states on the types of insurance plans offered.

Reference

Chen, W., & Page, T. F. (2018). Impact of Health Plan Deductibles and Health Insurance Marketplace Enrollment on Health Care Experiences. *Medical Care Research and Review*, *77*(5), 107755871881012. https://doi.org/10.1177/1077558718810129

Drake, C., Abraham, J. M., & McCullough, J. S. (2015). Rural Enrollment in the Federally Facilitated Marketplace. *The Journal of Rural Health*, *32*(3), 332–339. https://doi.org/10.1111/jrh.12149

Fisher, E. S., & Wennberg, J. E. (2003). Health care quality, geographic variations, and the challenge of supply-sensitive care. *Perspectives in biology and medicine*, *46*(1), 69-79.

Goldman, A. L., McCormick, D., Haas, J. S., & Sommers, B. D. (2018). Effects Of The ACA's Health Insurance Marketplaces On The Previously Uninsured: A Quasi-Experimental Analysis. *Health Affairs*, *37*(4), 591–599. https://doi.org/10.1377/hlthaff.2017.1390

Hass, Christopher (June 3, 2009). "President Obama Reiterates Support for Public Option and Health Insurance Exchange". *Obama for America*.

Health Insurance Exchange Public Use Files (exchange pufs). CMS. (n.d.). Retrieved December 6, 2022, from https://www.cms.gov/cciio/resources/data-resources/marketplace-puf

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning: With applications in R*. Springer.

McKillop, C. N., Waters, T. M., Kaplan, C. M., Kaplan, E. K., Thompson, M. P., & Graetz, I. (2018). Three years in – changing plan features in the U.S. health insurance marketplace. *BMC Health Services Research*, *18*(1). https://doi.org/10.1186/s12913-018-3198-3

National Conference of State Legislatures. (2019). State & Legislative Partisan Composition Table. Retrieved from https://www.ncsl.org/Portals/1/Documents/Elections/Legis_Control_2019_February%201st.pdf

Song, Y., Skinner, J., Bynum, J., Sutherland, J., Wennberg, J. E., & Fisher, E. S. (2010). Regional variations in diagnostic practices. *New England Journal of Medicine*, *363*(1), 45-53.

U.S. Centers for Medicare & Medicaid Services. (2019, March 25). *HEALTH INSURANCE EXCHANGES 2019 OPEN ENROLLMENT REPORT | CMS*. Www.cms.gov. https://www.cms.gov/newsroom/fact-sheets/health-insurance-exchanges-2019-open-enrollment-report#_ftn3