

RUPRI Center for Rural Health Policy Analysis

Technical Report

Brief No. 2015-4

March 2015

www.public-health.uiowa.edu/rupri

A Rural Taxonomy of Population and Health-Resource Characteristics

Xi Zhu, PhD; Keith J. Mueller, PhD; Thomas Vaughn, PhD; Fred Ullrich, BA

Introduction

It is well recognized that where we live affects our chances of having healthy lives as well as how we prevent and treat illness.^{1,2} Characteristics of communities and the health care delivery systems that serve them jointly determine how health services are delivered, accessed, financed, and sustained as well as the health outcomes of the population. Public policies and community strategies that aim to improve population health and health equity could be enhanced by an understanding of these community characteristics and by implementing targeted, place-based interventions that address contextual factors affecting access, quality, and cost of care.^{3,4} Meanwhile, innovations could be better disseminated and community partnerships could be formed to pursue collaborative strategies if there were standard methods for describing, comparing, and evaluating communities.^{5,6} Realization of such potentials relies on a systematic approach, or a taxonomy, for classifying and identifying *similar* communities and places based on a set of relevant characteristics.

This report summarizes the rationale and the methodology for developing an empirical taxonomy of rural places. The taxonomy will contribute to RUPRI's goals to inform rural communities and policy makers about the implications of health care system changes, to assist rural communities and providers adopt innovations and transition to a high performance health system, and ultimately to advance population health and well-being in rural communities using a place-based approach. To this end, we developed a taxonomy of rural places based on their relevant population and health-resource characteristics, including socio-demographics, economic indicators, health insurance coverage, and healthcare resources. Incorporating information related to both demand and supply sides of the health services market, this taxonomy provides a systematic tool for classifying and identifying similar rural communities and places.

Data and Measures

We used the most current data from multiple sources. Demographic and health insurance coverage data were obtained from the American Community Survey (ACS). The ACS, a product of the US Census Bureau, is an ongoing statistical sample of the U.S. population. It provides period estimates of population demographic, social, economic, and housing characteristics. The 2008-2012 version of ACS 5-year estimate data was used to provide data on the following variables: population aged 65 years and older, younger than 18 years, non-white, with household income less than 150% of the Federal Poverty Level (FPL), unemployed, uninsured, and public insured. All variables were converted to percentages using appropriate denominators (see Table 1 below).



Funded by the Federal Office of Rural Health Policy, Rapid Response Co-op Grant #U1G RH07633

RUPRI Center for Rural Health Policy Analysis, University of Iowa College of Public Health, Department of Health Management and Policy, 145 Riverside Dr., Iowa City, IA 52242-2007, (319) 384-3830
<http://www.public-health.uiowa.edu/rupri> E-mail: cph-rupri-inquiries@uiowa.edu

Data on health care providers (i.e., number of primary care physicians, medical specialists, non-physician practitioners, and dentists) were obtained from the September 2012 version of the National Provider Identifier (NPI) file.⁷ The NPI registry, established by the Centers for Medicare and Medicaid Services, contains address and identification information on all health care providers who can receive direct Medicare reimbursement. Provider type was identified using the Healthcare Provider Taxonomy Code⁸ and medical credential information (self-reported by providers) contained in the registry data. Primary care physicians included those indicating practices in Family Medicine, General Practice, Obstetrics and Gynecology, Internal Medicine, or Pediatrics. Medical specialist physicians included those indicating practices in one of 28 different medical specialties.⁹ Non-physician practitioners included Advanced Practice Midwife, Nurse Practitioner, and Physician Assistant. All provider counts were converted to per capita rates.

Hospital data (i.e., staffed beds and average daily census (ADC)) were obtained from the 2011 American Hospital Association (AHA) Annual Survey Data. Data on Medicare/Medicaid-certified nursing home beds were obtained from the Centers for Medicare & Medicaid Services (CMS) “Nursing Home Compare” data, which provides detailed information about all such facilities in the country. CMS “Nursing Home Compare” data from January 2013 was downloaded and used in the analysis.¹⁰ Hospital beds, ADC, and SNF beds were converted to per capita rates.

Table 1 summarizes the population and health-resources characteristics and the variables used to measure them.

Table 1. Variables Used in Developing the Taxonomy

Characteristic	Variables
Age	Percentage of population aged 65 years and older = population aged 65 years and older / total population Percentage of population younger than 18 years = population younger than 18 years / total population
Race/Ethnicity	Percentage of population that are non-white = population reported being non-white / total population
Income/Poverty	Percentage of population with household income less than 150% of the FPL = population with household income less than 150% of the FPL / population for whom poverty status is determined
Unemployment	Percentage of population that are unemployed = population unemployed in civilian labor force / population aged 16 years and older
Health Insurance	Percentage of population that are uninsured = population reported having no health insurance / civilian, non-institutionalized population Percentage of population that are publicly insured = population reported being publicly insured / civilian, non-institutionalized population
Hospital Facility	Staffed hospital beds per capita = number of staffed hospital beds / total population Average daily census per capita = ADC / total population
Certified Nursing Home	Medicare/Medicaid-certified nursing home beds per capita = number of SNF beds / total population
Providers	Primary care physicians per capita = number of primary care physicians / total population Medical specialists per capita = number of medical specialists / total population Non-physician practitioners per capita = number of non-physician practitioners / total population Dentists per capita = number of dentists / total population

Unit of Analysis

All data were originally obtained at the ZIP Code Tabulation Area (ZCTA) level and then aggregated to the Primary Care Service Area (PCSA) level as the unit of analysis. Developed and maintained by the Dartmouth Atlas of Health Care, PCSA is defined as “a ZIP code area with one or more primary care providers and any contiguous ZIP code areas whose Medicare populations seek the plurality of their primary care from those providers.”¹¹ Thus, a PCSA is the smallest geographic unit that can be used to meaningfully measure health care resources, utilization, and associated outcomes.¹² Since PCSAs reflect health care utilization patterns, for this project they are preferred to other geographic units of analysis (e.g., counties) that place arbitrary spatial limits on health services markets. The latest available version of PCSA data (2007) was downloaded from the Health Resources and Services Administration (HRSA) Health Workforce Analysis website.¹³ The compiled data set covered 6541 PCSAs in the U.S.

Analytical Approach

There are many PCSAs that have portions of their populations living in both rural and urban ZCTAs. In these cases, we calculated the percentages of PCSA population living in rural places, defined based on Rural Urban Commuting Area Codes (RUCAs) 4-10. After examining the distribution of the percentages, 4024 PCSAs with more than 25% of their populations living in rural ZCTAs were retained as rural PCSAs.

Following the guidelines suggested by the methodological literature on cluster analysis¹⁴⁻¹⁹, we conducted a series of analyses and sensitivity tests to identify clusters of rural places that share common characteristics within the cluster while are distinctive from places in other clusters. The analytical steps are summarized below:

1. Univariate analysis: We examined the variable distributions to identify outliers. Five PCSAs were identified as apparent outliers in the distribution of one or more variables, including Danville, PA; Hettinger, ND; West Lebanon, NH; Cooperstown, NY; and Grand Ronde, OR. These five PCSAs were excluded from the final sample to avoid unwarranted influence of extreme values. The final sample included 4019 PCSAs.
2. Principal component analysis: Many of the observed variables were highly correlated. To weight all variables in the clustering process while reducing the impact of multicollinearity, we conducted a principal component analysis with orthogonal varimax rotation to identify principal components that accounted for a maximal amount of variance of the observed variables. Four principal components were identified and the component scores were used in the cluster analysis.¹⁷
3. Preliminary cluster analysis: The average linkage method²⁰ incorporated in the SAS CLUSTER Procedure²¹ was used in this step to identify initial cluster solutions, select candidate number of clusters, obtain centroids of the initial clusters, and spot potential outliers.
4. Iterative cluster analysis and validation: The K-means method²² incorporated in the SAS FASTCLUS Procedure²³ was used as an iterative partitioning strategy to obtain and validate the final cluster solution.
5. Post-clustering analysis: We transformed the clustering variables (i.e., the four principal components) into the canonical dimensions (i.e., underlying differentiating functions) in a canonical discriminant analysis to examine what combinations of the clustering variables had produced maximal separation between clusters. We plotted clusters using both the

clustering variables and canonical dimensions to examine the location and separation of clusters. We then compared the distributions of the 4 clustering variables and the 14 original variables across clusters to develop substantive interpretations of the clusters, which resulted in a taxonomy of rural places.

6. Sensitivity tests: We conducted sensitivity tests by using alternative rural inclusion criteria (e.g., >50% rural population), alternative initial clustering algorithms (e.g., Ward's minimum variance method), different candidate numbers of clusters, and different starting points for iterative cluster analysis. The sensitivity tests suggested that the combination of the average linkage and *K*-means methods produced the best results (judged by the meaningfulness of differentiating factors and the clear separation of different clusters), which is consistent with previous research suggesting that these methods outperform other methods¹⁷. The analysis of a smaller sample with a more stringent inclusion criteria (i.e., >50% rural population) produced comparable results, which cross-validated the taxonomy.

Results

Principal Component Analysis

The principal component analysis revealed that four components, or key dimensions, accounted for 62.8% of the total variance observed in the fourteen original variables. Table 2 presents the rotated factor pattern, which shows a clear loading pattern with minimal cross-loading.

Table 2. Results of Principal Component Analysis

	Dimension1	Dimension2	Dimension3	Dimension4
Primary care physicians per capita	0.027	0.788	0.227	0.030
Medical specialists per capita	0.022	0.798	0.042	-0.016
Non-physician practitioners per capita	-0.059	0.566	0.319	0.114
Dentists per capita	0.120	0.675	-0.094	0.021
Staffed hospital beds per capita	0.014	0.269	0.827	-0.033
Average daily census per capita	0.051	0.200	0.823	-0.031
Certified nursing home beds per capita	0.152	-0.138	0.483	0.257
% pop. unemployed	0.646	-0.095	0.291	0.104
% pop. uninsured	0.655	0.094	0.026	0.243
% pop. publicly insured	0.867	0.071	0.022	-0.118
% pop. household income < 150% FPL	0.911	0.105	-0.026	0.028
% pop. non-white	0.699	-0.020	0.012	0.244
% pop. age > 65	0.164	-0.055	0.221	0.853
% pop. age < 18	-0.137	-0.187	0.092	-0.827

Based on the principal component analysis, we converted the 14 original variables into four key dimensions. Each dimension combined a set of highly correlated variables into a compound dimension score, which was used in the subsequent analyses. The first dimension is a linear combination of five observed variables (i.e., percentages of population that are unemployed, uninsured, publicly insured, with household income < 150% FPL, and non-white), weighted by the eigenvectors. These variables either directly indicate (e.g., insurance status) or are highly correlated with (e.g., socio-economic status such as employment, poverty, and minority) the ability to pay for care. Therefore, we labeled the first dimension *Economic Resources*. Because higher scores on these variables suggest less economic resources, we inversely coded the dimension scores such that a higher dimension score indicates more economic resources. The second dimension is a combination of four variables related to availability of health care providers; and thus is labeled *Provider Resources*. The third dimension is a combination of three variables related to healthcare facilities. The two bed count variables indicate the availability of beds while the ADC is an indicator of hospitals' operational capacities. Thus, we labeled the third dimension *Facility Resources*. The fourth dimension combines two age-related variables and is labeled as *Age Distribution*. A higher score on *Age Distribution* indicates a higher percentage of population aged 65 and older and a lower percentage of population younger than 18 years.

Because we used the orthogonal varimax rotation in the principal component analysis, the four dimensions are uncorrelated with one another.

Cluster Analysis

The preliminary cluster analysis with the average linkage method suggested that the optimal candidate number of clusters was 10. After examining the initial cluster solution, we identified no apparent outliers. Centroids of the initial clusters were used as the starting points for the iterative cluster analysis. The iterative partitioning procedure resulted in 10 clusters (observed overall r -squared = 0.633). Table 3 presents the results of a canonical discriminant analysis, showing the correlations between the clustering variables and the canonical dimensions. The canonical dimensions, also known as discriminant functions, are the linear combinations of the clustering variables that provide maximal separation between clusters. The correlations suggest that the first canonical dimension, which produces the most significant separation, is predominantly associated with *Facility Resources* ($r = 0.945$). The second and third canonical dimensions are highly associated with *Provider Resources* and *Economic Resources* with opposite combinations of these two variables. The fourth canonical dimension is predominantly associated with *Age distribution*.

Table 3. Total Canonical Structure among Clustering Variables

Variable	Can1	Can2	Can3	Can4
Economic Resources	0.198	0.726	0.635	-0.176
Provider Resources	0.255	-0.650	0.702	0.141
Facility Resources	0.945	0.011	-0.322	-0.053
Age Distribution	0.050	0.226	-0.004	0.973

Location and Separation of Clusters

Based on the canonical structure, we plotted the clusters and their centroids against the four variables in order from more differentiating to less differentiating variables. The results are shown in Figures 1-4. Combining these results, we identified Clusters 1-3 as a distinctive group with more facility resources than other clusters. Clusters 4 and 5 are distinctive from others because they have more provider resources. Clusters 6-10 are differentiated from Clusters 1-5 and among themselves based on the combinations of economic resources and age distribution. More specifically, Clusters 6 & 7 have higher economic resources and Cluster 8 has higher age distribution scores (i.e., higher percentages for age 65 and older and lower percentages for under age 18).

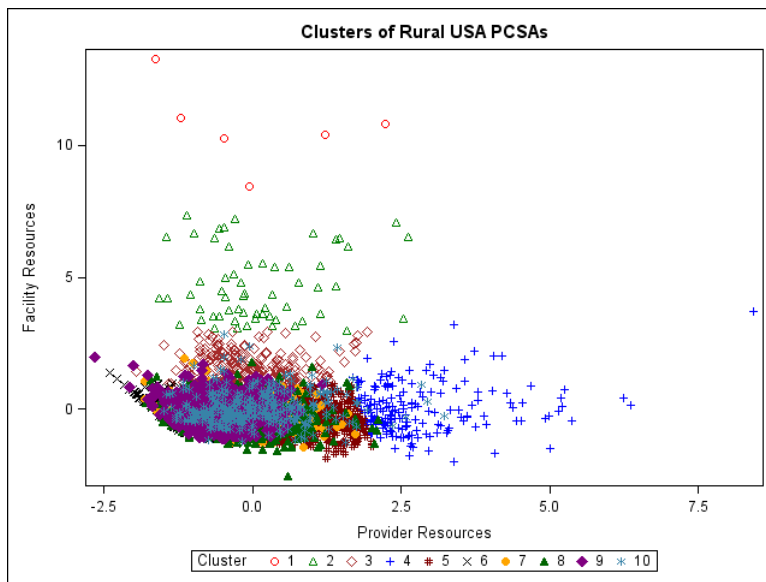


Figure 1. Facility and Provider Resources of Rural PCSA Clusters

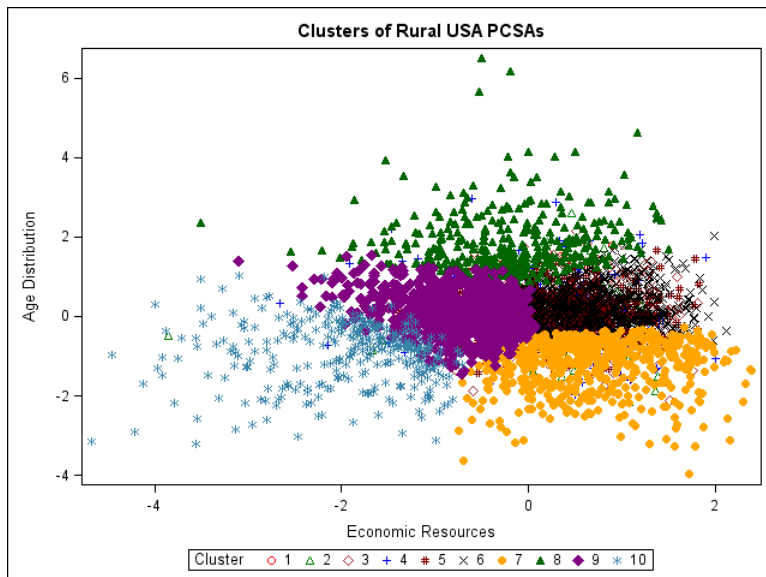


Figure 2. Economic Resources and Age Distribution of Rural PCSA Clusters

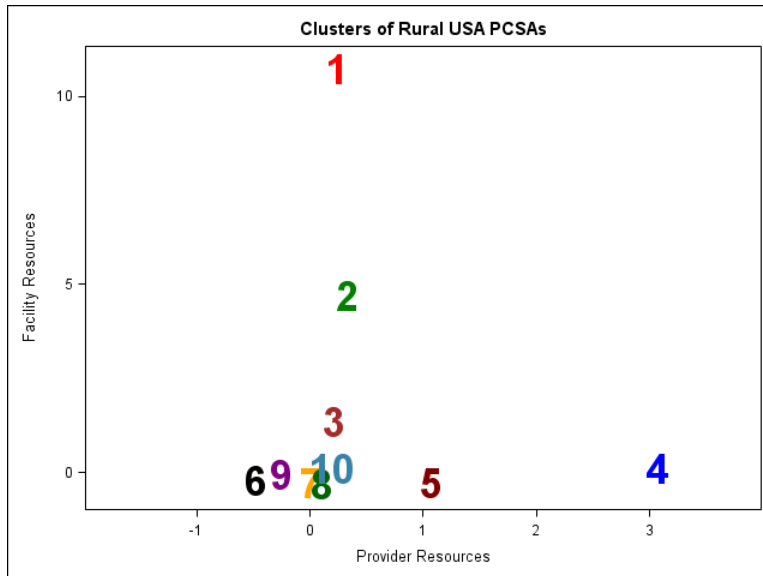


Figure 3. Facility and Provider Resources of Cluster Centroids

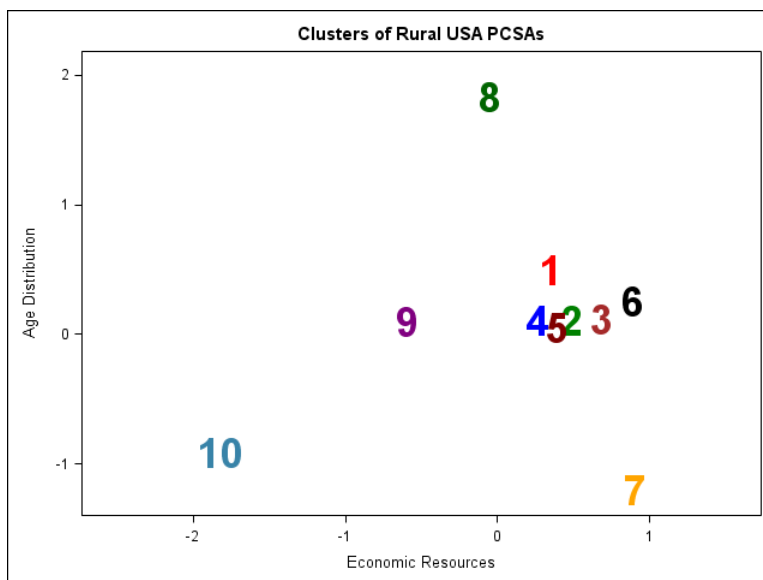


Figure 4. Economic Resources and Age Distribution of Cluster Centroids

The Taxonomy

The number of PCSAs in each cluster and the differentiating characteristics are presented in Table 4. Because the clustering variables were standardized (i.e., mean=0 and standard deviation=1), a score of zero on one dimension (e.g., facility resources) indicates that a PCSA is average on that dimension (e.g., having average facility resources per capita). Positive scores indicate resource availability or age distribution higher than the average, and negative scores indicate resource availability or age distribution lower than the average. Following these interpretations, we characterized clusters based on the distribution of their members along the four clustering dimensions. Empirically, we described a cluster as *average* on one dimension if the cluster's members scattered around zero on this dimension and the cluster mean approximated zero. We described a cluster as *high* or *low* on a dimension when all, or in a few

cases the majority, of the cluster's members had scores larger or smaller than zero and the absolute value of cluster mean ranged between 0.7 and 2. We described a cluster as *very high* or *extremely high* on a dimension when all members of the cluster had scores larger than zero (empirically larger than 1.6) and the cluster mean ranged between 2 and 5 or above 5.

Table 4. A Rural Taxonomy of Population and Health-Resource Characteristics

Cluster	N	Facility Resources	Provider Resources	Economic Resources	Age Distribution
1	6	Extremely High	Average	Average	Average
2	59	Very High	Average	Average	Average
3	318	High	Average	Average	Average
4	179	Average	Very High	Average	Average
5	686	Average	High	Average	Average
6	743	Average	Low	High	Average
7	574	Average	Average	High	Low
8	364	Average	Average	Average	High
9	771	Average	Average	Low	Average
10	319	Average	Average	Low	Low

As noted above, the four dimensions used in the clustering analysis are uncorrelated with one another. Consequently, the clustering results and the taxonomy show that the ten clusters are differentiated from others by one or two defining characteristics, which is a desirable outcome. For a particular cluster, its undifferentiating dimensions are described as average because the member PCSAs of the cluster are distributed around the population means on these dimensions. Below we summarize the defining characteristic(s) of each cluster in the taxonomy:

- Cluster 1 has extremely high facility resources with the overall facility (i.e., staffed hospital beds, ACD, and/or SNF beds) per capita rate ranging between 8.4 and 13.3 standard deviations above the population mean;
- Cluster 2 has very high facility resources with the overall facility per capita rate ranging between 3.0 and 7.4 standard deviations above the population mean;
- Cluster 3 has high facility resources with the overall facility per capita rate ranging between 0.4 and 2.9 standard deviations above the population mean;
- Cluster 4 has very high provider resources with the overall provider (i.e., primary care physicians, specialists, non-physician practitioners, and/or dentists) per capita rate ranging between 1.7 and 8.4 standard deviations above the population mean;
- Cluster 5 has high provider resources with the overall provider per capita rate ranging between -0.1 and 2.0 standard deviations above the population mean;
- Cluster 6 has low provider resources and high economic resources with the overall provider per capita rate ranging between -2.4 and 0.4 and the overall economic resources (i.e., %s of employed, insured, non-publicly insured, with household income > 150% FPL, and/or white) score ranging between -0.2 and 2.1;

- Cluster 7 has high economic resources with the overall economic resources score ranging between -0.8 and 2.4, and a younger population with a composite age distribution score ranging between -4.0 and -0.3;
- Cluster 8 is defined by an older population with a composite age distribution score ranging between 0.7 and 6.5;
- Cluster 9 has low economic resources with the overall economic resources score ranging between -3.1 and 0.1;
- Cluster 10 has low economic resources with the overall economic resources score ranging between -4.7 and -0.7, and a younger population with a composite age distribution score ranging between -3.1 and 1.0.

Figure 5 shows that the clusters are clearly separated from one another based on the dimensions that define them (i.e., Facility Resources for Clusters 1-3, Provider Resources for Clusters 4 and 5, and Economic Resources and Age Distribution for Clusters 6-10) as shown in the upper left and the lower right graphs, but are almost indistinguishable based on the other dimensions as shown in the upper right and the lower left graphs. The results suggest that the taxonomy has several desirable features including interpretability and clear separation.

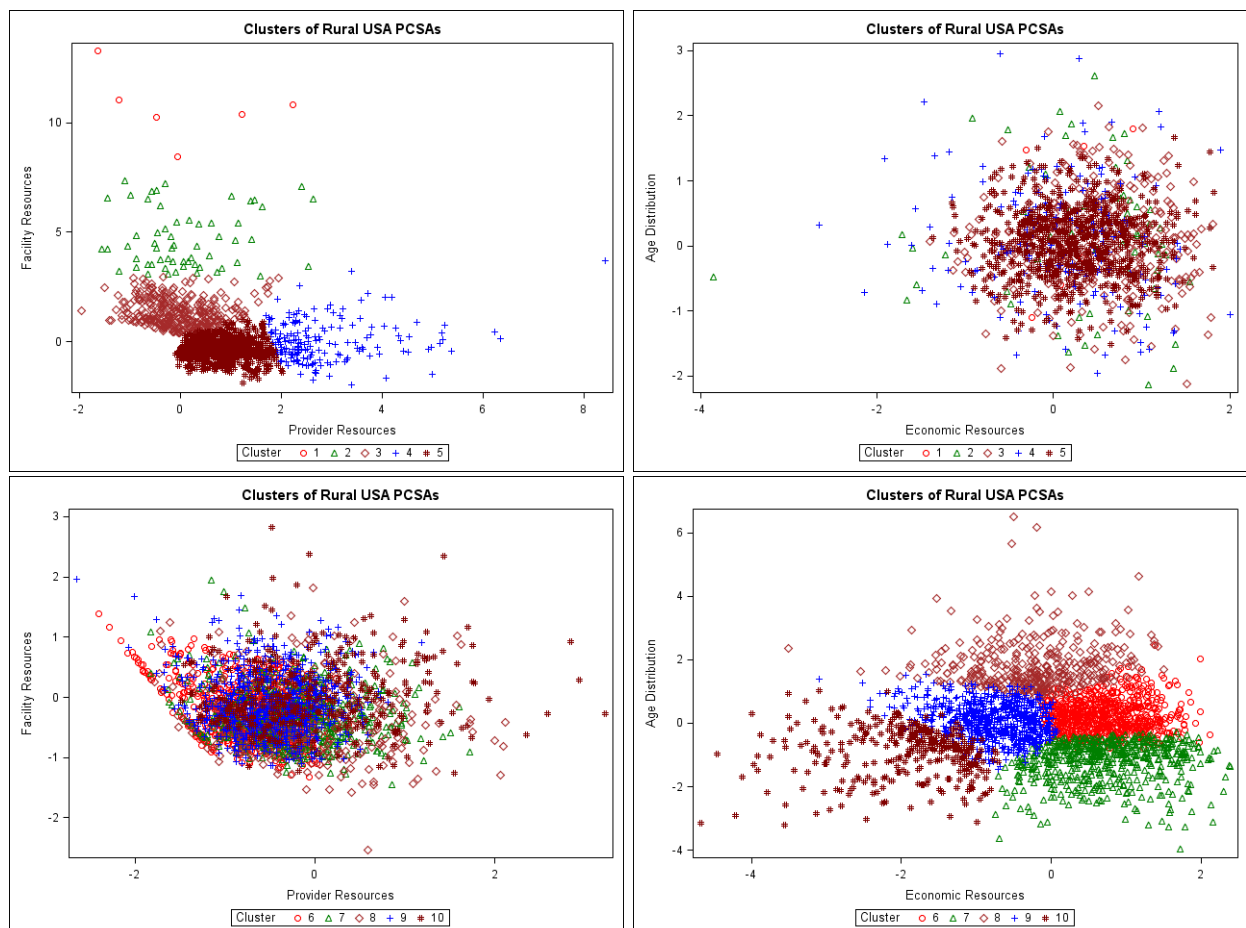


Figure 5. Separation of Clusters

Distribution of Clustering Variables

Figures 6-9 display the distributions of the four dimension scores across different clusters. The distributions confirmed the defining characteristics of the taxonomy, especially for Clusters 1-5. Additionally, it shows that although Clusters 6 and 7 have higher Economic Resources compared to Clusters 8-10, their scores are not distinctively higher than those of Clusters 1-5 (i.e., the means are higher, but the interquartile ranges are overlapped). This is because the clustering methods consider the combination of all four dimensions in differentiating places. The first two dimensions (Facility Resources and Provider Resources) are more differentiating and clearly distinguish Clusters 1-5 from the rest of clusters. It is worth noting that the ranges of Economic Resources for Clusters 1-5 are much wider than those for Clusters 6 and 7, and the interquartile ranges contain zero, which suggests that Economic Resources is not the defining characteristic for Clusters 1-5. Lastly, consistent with the taxonomy, Cluster 8 stands out as having higher scores on Age Distribution and Clusters 7 and 10 are low on this dimension.

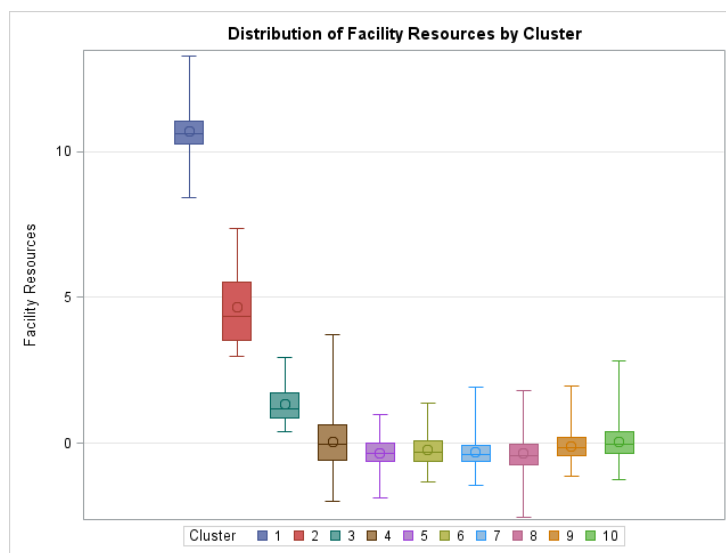


Figure 6. Distribution of Facility Resources by Cluster

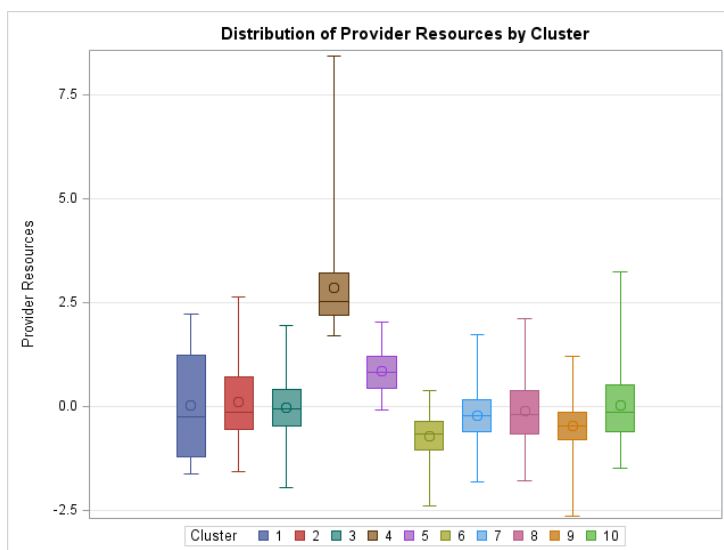


Figure 7. Distribution of Provider Resources by Cluster

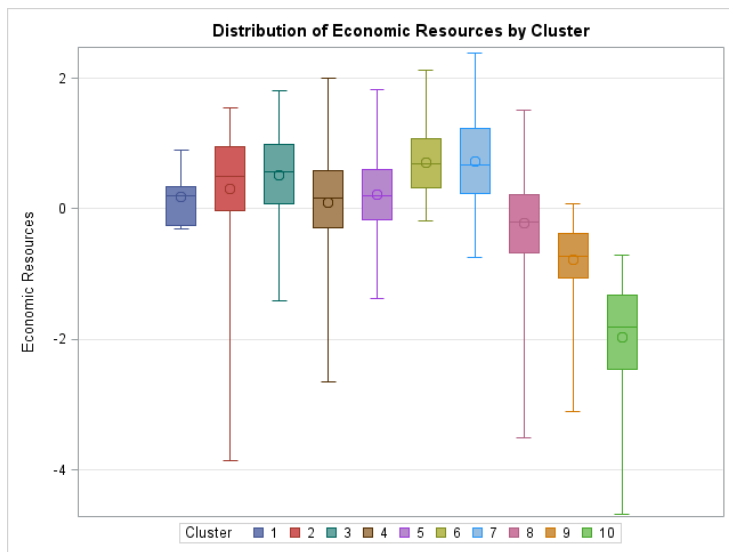


Figure 8. Distribution of Economic Resources by Cluster

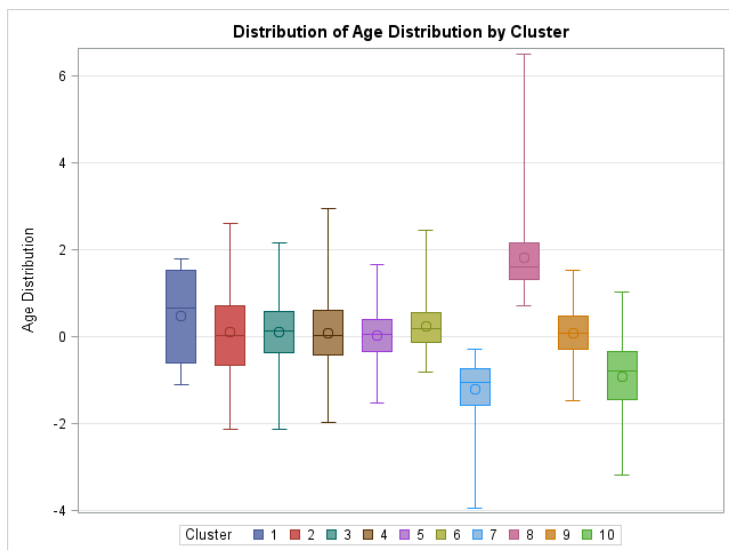


Figure 9. Distribution of Age Distribution by Cluster

Sample State Maps Informed by the Taxonomy

Figures 10-13 present maps of PCSAs for four sample states (Iowa, Montana, North Carolina, and Pennsylvania) where PCSAs are color-coded with their designated cluster numbers in the taxonomy.

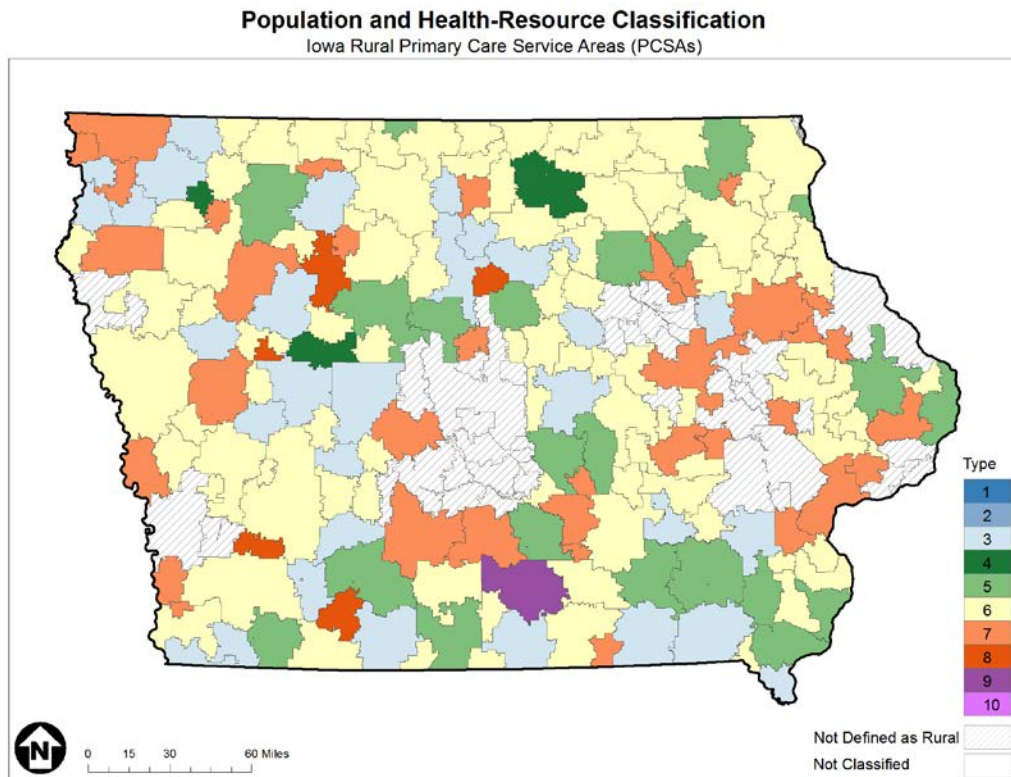


Figure 10. Iowa PCSAs with Designated Classification

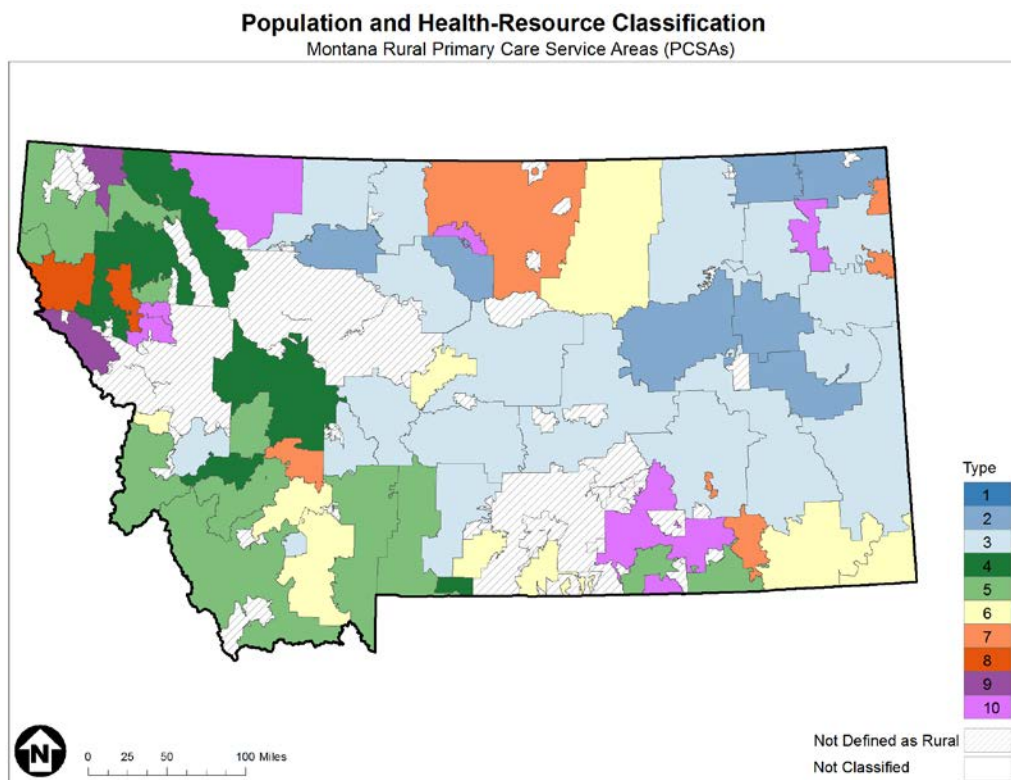


Figure 11. Montana PCSAs with Designated Classification

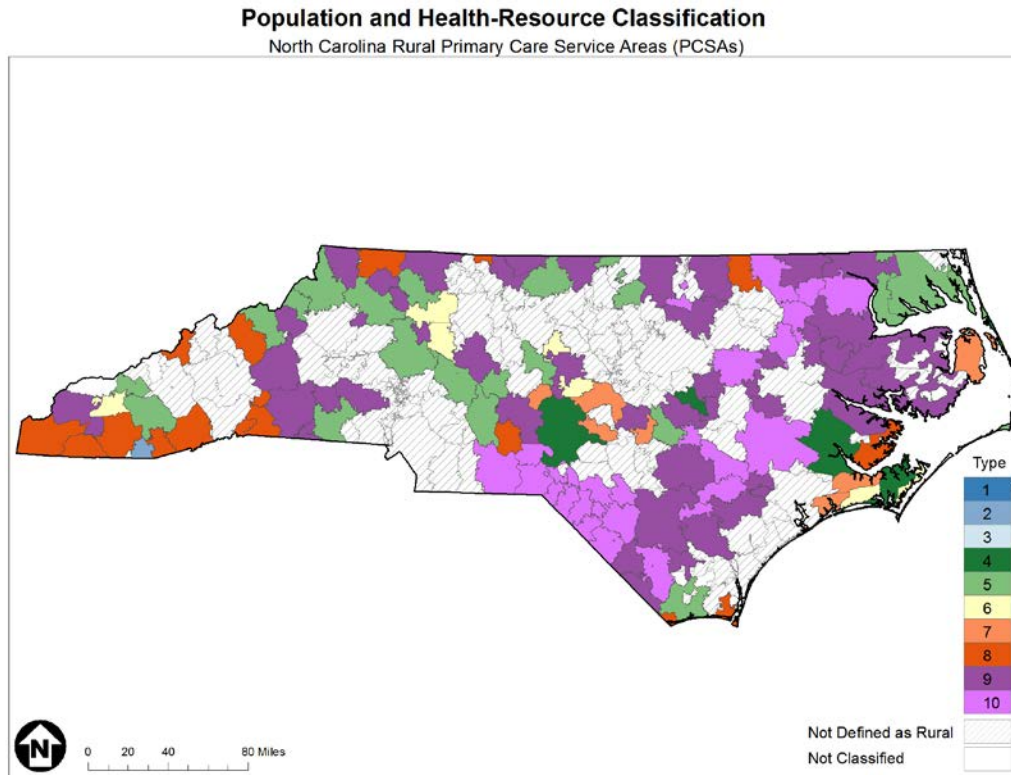


Figure 12. North Carolina PCSAs with Designated Classification

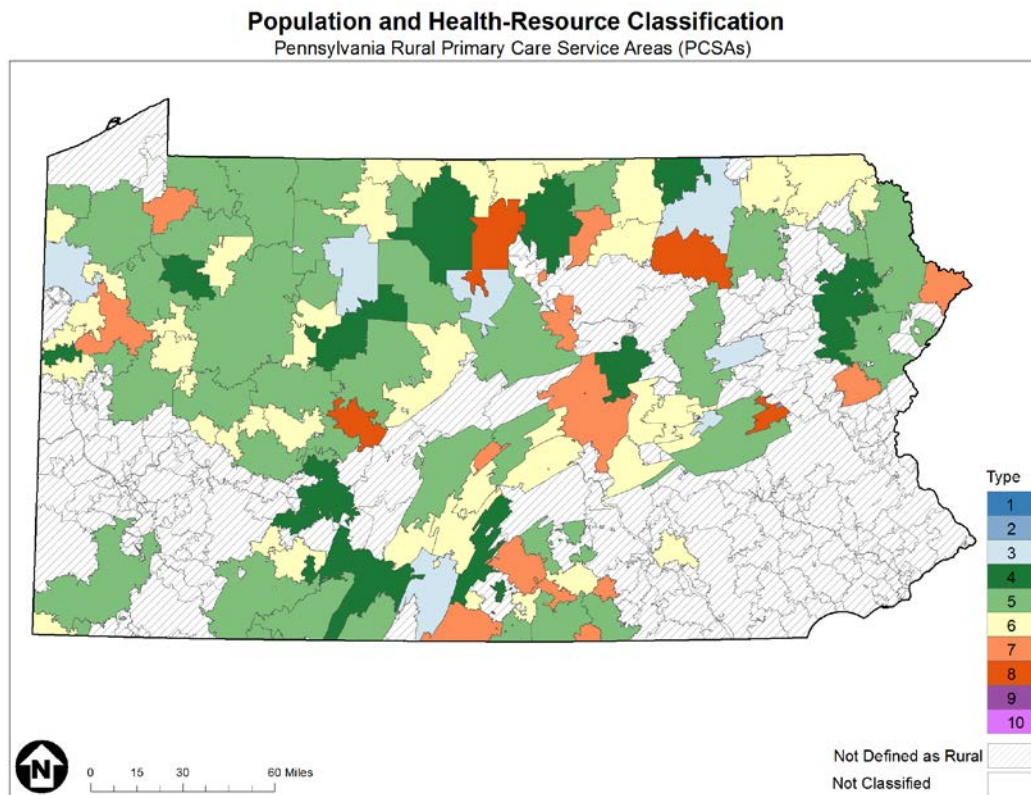


Figure 13. Pennsylvania PCSAs with Designated Classification

Discussion

In this report, we described the rationale and methods for developing a taxonomy of rural places based on population and health-resource characteristics. Our analyses built on characteristics of both rural communities and health delivery systems to profile how community and healthcare resources are *currently* organized at the PCSA level. The taxonomy that we have developed has several desirable features: (1) using PCSA geographies provides a “cleaner” definition of community characteristics and health resources as it allows the self-identification of “community” based on the healthcare-seeking behavior of the population in an area – such behavior rarely respects political boundaries (e.g. counties); (2) a reasonably small number of types accounted for a large amount of variation in community characteristics; (3) all but one type had a substantial number of PCSAs, indicating that the taxonomy was not heavily influenced by a few outliers or outlier groups; (4) all types of PCSAs were clearly separated from one another; and (5) while taking into account many characteristics (14 original variables and 4 key dimensions), the 10 types of PCSAs in the empirical taxonomy were mostly distinct from others on one or two defining dimensions.

This taxonomy of rural PCSAs can be used to inform rural health policy making; help rural communities develop strategies, adopt innovations, and form learning collaboratives; and extend health services research by incorporating typological characteristics of places (i.e., the combination of characteristics that differentiate a place) in the investigation of access, spending, and outcomes of health care.

For policy makers and analysts, the taxonomy can be used to simulate the effects of policy changes on rural communities. For instance, using the data for a rural place as the “base case,” simulations could explore the impact of policy changes on the rural community. The examples below explore the applications of the analysis:

- Example 1: one of the variables identified above is the insurance status of the residents in the location. Under the ACA, this insurance status could be assumed to change, and the data produced here could be used to analyze impacts on the rural health system.
- Example 2: another variable identified above is the demographic characteristics of the population base of the area (e.g., age distribution). As the country’s population continues to age, it will be important to explore the impact of a shift in the population of an area towards having more elders over age 65, thus leading to more demand for Medicare and Medicaid services.

For rural communities, the taxonomy can be used to assess the community’s own profile, identify similar communities, and develop strategies using a comparative framework. One challenge for rural communities to participate in national demonstration and pilot programs for health system innovation was the lack of a tool for assessing whether an innovation was applicable in their “unique” context. The taxonomy provides a baseline description of a community’s profile regarding its essential demographic, socio-economic, insurance, and health-resource conditions in comparison to other rural communities. Community leaders could search for meaningful comparisons among communities by: 1) identifying communities from the same cluster in the taxonomy and 2) considering other characteristics relevant for health system innovation such as those related to market conditions (e.g., numbers of clinics and other health care organizations in the area, competition among clinics and providers), the system (e.g., whether different parts

of care delivery system are integrated), geography (e.g., distance to tertiary care, spread of the population), and culture (e.g., care-seeking patterns of the community members). Building on such comparisons, rural communities could adopt innovations that are successfully implemented in similar communities or develop learning collaborative with such communities.

Limitations

Despite its potential usefulness for informing rural health policy, community/system development, and research, the taxonomy developed here has several important limitations that deserve special attention. First and most notably, given the purpose of this analysis and the nature of the methodology (especially those of cluster analysis), we focused on limited numbers of social-demographic, insurance, and health care resource variables in differentiating rural communities and used the per capita rates as standard measures to account for the impact of population size. Due to data and methodological constraints, we omitted several important factors related to the delivery, access, finance, and sustainment of health care in rural communities, which may undermine the utility of the taxonomy. These factors include:

- Geography-related factors: The size of the PCSA, population density, the distribution of health care providers and facilities *within* the PCSA, travel distance to care, adjacent areas and their characteristics;
- Market- and system-related factors: The relationship between different providers and health care organizations, system affiliation and network participation, the experience with care coordination across the continuum;
- Facility-related factors: The availability of different types of facilities (e.g., general hospitals, CAHs, FQHCs);
- Culture-related factors: The care-seeking patterns of community members, cultural norms related to care provision (e.g., acceptance of palliative care).

When used to assist particular rural communities to develop place-based strategies or interventions, the taxonomy can *only* provide a baseline profile of the communities and must be combined with more detailed analyses of other community characteristics to draw valid inferences on which communities are in fact similar to one another. For instance, a frontier community consisting of sparsely populated areas that are isolated from population centers may be classified into the same cluster as a small, relatively concentrated rural community adjacent to a micropolitan area due to a unique combination of demographics and resources. However, care delivery and access in those communities may differ significantly due to the impact of commute time. As a result, these communities may need different strategies or interventions to effectively improve health access and outcomes for their community members. Therefore, the additional characteristics must be incorporated into policy and strategic considerations.

Second, we opted to use the PCSA as the unit for classification because it is the smallest geographic definition based on health services utilization. However, PCSAs may not be a natural unit for civic planning, policy making, or intervention. Two alternatives would be to use a different service area definition such as hospital service areas or a geopolitical definition such as counties. The former retains the disadvantage of incongruence with civic boundaries while creating a larger geography that loses some of the relevance to small rural places. The latter would conform to civic boundaries, but lose relevance to patterns of health services utilization. Future research is needed to cross-validate the classification results based on different geographic definitions and to test the utility of different classification systems in explaining

current health care outcomes and in facilitating future development of place-based policy and interventions.

References and Endnotes

1. Evans RG, Barer ML, Marmor TR. *Why are some people healthy and others not? The determinants of health of populations*. Hawthorne, NY: Aldine de Gruyter; 1994.
2. Marmot M, Friel S, Bell R, Houweling TAJ, Taylor S. Closing the gap in a generation: Health equity through action on the social determinants of health. *The Lancet*. 2008;372:1661-1669.
3. Acevedo-Garcia D, Osypuk TL, McArdle N, Williams DR. Toward A Policy-Relevant Analysis Of Geographic And Racial/Ethnic Disparities In Child Health. *Health Affairs*. March 1, 2008 2008;27(2):321-333.
4. Amaro H. The action is upstream: Place-based approaches for achieving population health and health equity. *American Journal of Public Health*. 2014;104(6):964-964.
5. Roussos ST, Fawcett SB. A Review of Collaborative Partnerships as a Strategy for Improving Community Health. *Annual Review of Public Health*. 2000/05/01 2000;21(1):369-402.
6. Greenhalgh T, Robert G, Bate P, Macfarlane F, Kyriakidou O. *Diffusion of innovations in health service organisations: a systematic literature review*. Oxford, UK: BMJ Books/Blackwell; 2008.
7. http://nppes.viva-it.com/NPI_Files.html
8. <http://www.cms.gov/Medicare/Provider-Enrollment-and-Certification/MedicareProviderSupEnroll/Taxonomy.html>
9. Allergy and Immunology, Anesthesiology, Colon and Rectal Surgery, Dermatology, Electrodiagnostic Medicine, Emergency Medicine, Hospitalist, 'Other' Internal medicine, Neurological Surgery, Neuromusculoskeletal Medicine & OMM, Neuromusculoskeletal Sports Medicine, Nuclear Medicine, Ophthalmology, Oral and Maxillofacial Surgery, Orthopedic Surgery, Otolaryngology, Pain Medicine, 'Other' Pediatrics, Phlebology, Physical Medicine & Rehabilitation, Plastic Surgery, Preventive Medicine, Psychiatry and Neurology, Radiology, Surgery, Thoracic Surgery, Transplant Surgery, Urology
10. <http://www.medicare.gov/download/DownloadddbInterim.asp>
11. <http://www.hrsa.gov/advisorycommittees/shortage/Meetings/20110118/MeetingMaterials/primarycare.pdf>
12. Goodman DC, Mick SS, Bott D, et al. Primary care service areas: A new tool for the evaluation of primary care services. *Health Services Research*. 2003;38(1p1):287-309.
13. <http://bhpr.hrsa.gov/healthworkforce/data/primarycareserviceareas/index.html>
14. Ward JH, Jr. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*. 1963;58(301):236-244.
15. Anderberg MR. *Cluster analysis for applications*. New York: Academic Press; 1973.
16. Blashfield RK, Aldenderfer MS. The literature on cluster analysis. *Multivariate Behavioral Research*. 1978;13(3):271-295.
17. Punj G, Stewart DW. Cluster Analysis in Marketing Research: Review and Suggestions for Application. *Journal of Marketing Research*. 1983;20(2):134-148.
18. Ketchen DJ, Shook CL. The application of cluster analysis in strategic management research: An analysis and critique. *Strategic Management Journal*. 1996;17(6):441-458.
19. Kaufman L, Rousseeuw PJ. *Finding groups in data: An introduction to cluster analysis*. Vol 344. Hoboken, NJ: John Wiley & Sons; 2009.
20. Sneath PH, Sokal RR. *Numerical taxonomy*. San Francisco, CA: W. H. Freeman; 1973.
21. <https://support.sas.com/documentation/cdl/en/statugcluster/61777/PDF/default/statugcluster.pdf>
22. Howard N, Harris B. A hierarchical grouping routine, IBM 360/65 FORTRAN IV program. *University of Pennsylvania Computer Center*. 1966.
23. <https://support.sas.com/documentation/cdl/en/statugcluster/61777/PDF/default/statugcluster.pdf>