

Determining county-level counterfactuals for evaluation of population health interventions: A novel application of *K*-means cluster analysis

Keywords: evaluation, methods, population health

Authors:

Kelly L. Strutz, PhD, MPH, Assistant Professor¹

Zhehui Luo, PhD, Associate Professor²

Jennifer E. Raffo, MA, Project Coordinator¹

Cristian I. Meghea, PhD, Associate Professor¹

Peggy Vander Meulen, MSN, RN, Program Director^{3,4}

Lee Anne Roman, MSN, PhD, Professor¹

1. Department of Obstetrics, Gynecology and Reproductive Biology, Michigan State University College of Human Medicine, East Lansing and Grand Rapids, MI, USA
2. Department of Epidemiology and Biostatistics, Michigan State University College of Human Medicine, East Lansing, MI, USA
3. Strong Beginnings Federal Healthy Start Program, Spectrum Health, Grand Rapids, MI, USA
4. Healthier Communities, Spectrum Health, Grand Rapids, MI, USA

Corresponding Author: Kelly L. Strutz, PhD, MPH, Department of Obstetrics, Gynecology and Reproductive Biology, Michigan State University College of Human Medicine, Secchia Center, 15 Michigan St NE, Grand Rapids, MI 49503. 616-234-0982. strutzk1@msu.edu.

Total word count: 2296

Number of tables and figures: 5 total (3 tables, 2 figures)

Abstract

Objectives: Evaluating population health initiatives at the community level necessitates valid counterfactual communities, which includes having similar complexity with respect to population composition, healthcare access, and health determinants. Estimating appropriate county counterfactuals is challenging in states with large inter-county variation. We present and discuss an application of *K*-means cluster analysis for determining county-level counterfactuals in an evaluation of a county perinatal system of care for Medicaid-insured pregnant women.

Materials and Methods: Counties were described using indicators from the American Community Survey, Area Health Resources Files, University of Wisconsin Population Health Institute County Health Rankings, and vital records for Michigan Medicaid-insured births for the year intervention began (or the closest available year). We ran analyses of 1,000 iterations with random starting cluster values for each of a range of number of clusters from 3 to 10 and used standard variability and reliability measures to identify the optimal number of clusters.

Results: One county was grouped with the intervention county in all solutions for all iterations and thus considered most valid for 1:1 population county comparisons. Two additional counties were frequently grouped with the intervention county. However, no county was ideal for all subpopulation analyses.

Practice Implications: Although the *K*-means method was successful at identifying a comparison county, concerning intervention-comparison differences remained. This limitation of the method may be specific to this county and the constraints of a within-state study. This method could potentially be more useful when applied to other counties in and outside of Michigan.

Introduction

Complex, multifactorial public health problems necessitate population-wide solutions, and thus community-level interventions are frequently utilized. The gold standard for evaluating community-level interventions, such as those at the county level, is the group-randomized trial.¹ However, this is often not feasible or appropriate with a small number of communities, particularly for community-based participatory research and other scenarios where the intervention is community- or stakeholder-initiated, or system of care interventions that depend on coordinating local resources. It is not clear how best to evaluate non-randomly assigned community interventions in this scenario.²⁻⁴ As in individual-level quasi-experimental analyses, the challenge is to identify an appropriate counterfactual group or communities that should be as similar as possible to the intervention community/ies to approximate group randomization,^{3,5} but little guidance exists on selection criteria or methods.

We conducted a demonstration project from 2009-2015 to determine whether a county-level perinatal system of care enhancement improved population outcomes for Medicaid-insured pregnant women and their infants in Kent County, Michigan, a mixed urban/rural county containing the second-largest city in the state. Because our intervention targeted the entire county, the ideal counterfactual community or communities would have the same sociodemographic characteristics and level of complexity in service delivery as the intervention county. Complexity is influenced by many factors, including the numbers of residents, health care agencies, providers, and enhanced prenatal care programs.

Because we wanted to account for the broader context and administrative features of an entire county, we explored a clustering method, *K*-means cluster analysis, to identify one or more comparison counties in Michigan. *K*-means cluster analysis has been used increasingly in the public health and health services research literature to create clusters of exposures or individuals, most frequently for sources of air pollution⁶ and for health care population segmentation.⁷ Where it has been used previously for county-level clustering, it was employed to create clusters of similar counties to compare among each other.^{8,9}

The objective of this study was to present and discuss an application of *K*-means cluster analysis for determining county-level counterfactuals for evaluation of a county perinatal system of care for Medicaid-insured pregnant women. To our knowledge, this is the first use of this method to find the most similar counterfactual county.

Materials and Methods

Data

There are 83 counties in the state of Michigan. County-level features were selected from three publicly available data sources – the American Community Survey,¹⁰ the Area Health Resources Files,¹¹ and the University of Wisconsin Population Health Institute County Health Rankings.¹² The fourth data source was a limited dataset of Michigan resident live birth records for Medicaid-insured births only (hereafter called Michigan Medicaid vital records (MMVR)), including county identifiers. The file came from the Michigan Department of Health and Human Services (MDHHS) Division for Vital Records and Health Statistics, and was retrieved by an honest broker from the MDHHS Health Services Data Warehouse.

The following categories of features were considered: population composition, social and economic factors, health care access, health outcomes, health behaviors, and physical environment. We selected a total of 35 features across these broad categories, some chosen from Rettenmaier and Wang¹³ and others based on the evaluation team’s experience with the Michigan Medicaid Maternal and Infant Health Program and the Strong Beginnings federal Healthy Start program.¹⁴⁻¹⁶ These represented a mixture of compositional (i.e., those aggregated from individual-level characteristics) and contextual (i.e., those pertaining to the county environment) measures.¹⁷ Whenever possible, we used year 2009 data because it was the year when system implementation of pregnancy-centered service improvement began. For data elements without 2009 information, we used as proxies, in priority order, 2010, 2008, or 2011 data. Four features were missing data for certain counties (n ranged from 14-35) and mean values were imputed. A description of the 35 features utilized is in Table 1 along with the data source, year, mean, and standard deviation.

[insert Table 1 here]

Analysis

K-means cluster analyses were used to identify the comparison county or counties that would be classified in the same cluster as the intervention county in the most optimal model. Cluster analysis, also called data segmentation, groups or segments a collection of objects (counties in our study) into subgroups or clusters such that those within the same cluster are more similar to each other than objects (counties) assigned to different clusters.¹⁸ Counties are described by a set of characteristics (features, measurements) that are pre-selected based on the purpose of evaluations; in our case, relevant to

evaluations of pregnancy-centered health programs. Counties in the same cluster may have different degrees of similarity or dissimilarity to each other and to a specific county (Kent in our study). Among the most widely used clustering methods is the *K*-means (or median) clustering for which the within-cluster variation (WCV) is as small as possible.^{6,7}

We tested solutions for 3 to 10 clusters for 1,000 iterations each with randomly assigned starting cluster values. To identify the best solution, we used scree plots to visualize a kink in the curve generated from the within sum of squares (WSS) or its logarithm ($\log[\text{WSS}]$) for all cluster solutions, i.e., the point at which reduction in WSS or $\log\text{WSS}$ is not appreciable enough to warrant increasing the number of clusters.¹⁹ The WSS can be thought of as the error sum of squares in regression analysis where the total sum of squares is defined when all objects are in one cluster. Another criterion used was the η^2 coefficient, $\eta^2(k) = 1 - \frac{\text{WSS}(k)}{\text{WSS}(1)}$, which is between 0 and 1 with higher values indicating better clustering. Finally, we also transformed the WSS to get the proportional reduction of error (PRE) as well to identify the number of clusters that leads to the largest reduction in PRE.

Nineteen of the 35 features are proportions which have a natural range (0 and 100); and the other features may have large variation. It was conceivable that if we standardized these features some natural variation, e.g., the size of the population in each county, would be masked. In *K*-means analyses, if all variables are standardized then clustering based on correlation (similarity) is equivalent to that based on squared distance (dissimilarity).²⁰ Therefore, we ran all analyses twice, with and without standardizing all variables. Stata v.15 (StataCorp, College Station, TX) was used for all analyses. The evaluation project was reviewed by the Michigan State University, MDHHS, and Spectrum Health institutional review boards and determined not human subject research.

Results

Figure 1 shows that the optimal number of clusters when features were not standardized was 4, where WSS could be seen at a kink point and the reduction in PRE was the largest. On the other hand, Figure 2 shows that when features were standardized there was no clear optimal solution of the number of clusters.

[insert Figure 1 and Figure 2 here]

Using the unstandardized features, the number of times that a county was grouped in the same cluster as Kent County out of 1,000 iterations when the number of clusters varied from 3 to 6, the range around the optimal number of 4, is shown in Table 2. Three counties were grouped with Kent under

these criteria. Macomb County was grouped in the same cluster with Kent County in all iterations of all solutions, while Oakland County clustered with Kent in 979-1000 iterations and Genesee County clustered in 127-1000 iterations depending on number of clusters. In contrast, when we standardized all features, no county was grouped with Kent in a majority of iterations (results available upon request). Given the lack of an optimal solution under standardization, and that the counties grouped together were not similar, we opted to use the unstandardized variables.

[insert Table 2 here]

Table 3 lists the values of the 35 features for Kent, Macomb, Oakland, and Genesee counties. In examining the values individually, Macomb was within 2 percentage points of Kent for 10 of 19 percentage-based features, with nearly identical median household income and numbers of Medicaid-eligible women and minors. Oakland and Genesee were within 2 percentage points of Kent for 5 of 19 percentage-based features each, also both with similar median household incomes. There was broad variation in access to care features across counties. However, there were two related features for which all of the three comparison counties had values considerably smaller than Kent's: the two Latinx population features, "percent Latina females ages 15-54" (Kent 6.7, Macomb 1.5, Oakland 2.4, Genesee 2.0) and "percent Latinx aged 25+ with less than a high school diploma" (Kent 46.3, Macomb 22.4, Oakland 27.0, Genesee 16.9).

[insert Table 3 here]

Discussion

K-means clustering achieved identification of one to three comparison counties when pre-selected features were unstandardized in our study. Yet, the features that differed between the intervention county and other similar counties were concerning. In particular, the intervention county had more Latina women, and more with less than a high school education, than the best matched county by the *K*-means analysis. We were also cautious about differences in county maternal and infant health programming and infrastructure that were not reflected in the data sources. For example, one of the selected counties includes the city of Flint that would receive an influx of maternal and child health resources associated with the water crisis during the study period. Ultimately, we were able to achieve better covariate balance by selecting individual women from across the state (outside the intervention county) and using propensity score-weighted analyses, matched exactly on race and using more granular census tract-level and block group-level covariates.

These results are consistent with the findings of Wallace et al.,⁸ who used *K*-means clustering with national U.S. data to create clusters of counties similar on sociodemographic features. They noted considerable within-state heterogeneity, such that most states (including Michigan) contained counties from many different clusters. Of note, the three comparison counties that we identified were in the same cluster as our intervention county in Wallace’s schema as well. However, they also identified that the counties that were themselves sociodemographically heterogeneous would benefit from more granular, “subcounty” data, which we found applicable to our diverse intervention county.

Our findings are also congruent with recent discussion of how machine learning algorithms can reproduce biases against minoritized groups.²¹⁻²³ While previous literature has focused on racial/ethnic disparities in predictive algorithms, our analysis demonstrates how an ethnic subgroup can be obscured by a clustering method. Thus, the same cautions that should be extended to ensure equity in predictive modeling should be applied to cluster analyses as well.

General limitations of the *K*-means method to select counterfactual communities include the following. First, the results are only as good as the chosen features. In our example, features were selected through discussion by the research team and community partners with substantive knowledge of factors influencing perinatal health in the community. We were able to capture a wide range of important indicators of complexity, including health care determinants and social determinants of maternal and infant health outcomes, triangulated across four different data sources. That said, our selection was likely not the optimal set, and it is probable that no one optimal set exists. A more data-driven approach could be applied to feature selection as well.⁸ Relatedly, given our interest in the Latinx subpopulation, adding more measures specific to this subpopulation would weight their importance more heavily. However, this would not address the issue that there aren’t appreciable Latinx subpopulations in the counties whose other features are most similar to our intervention county, and would have led to no better solution.

Second, this method involves the assumption that if a counterfactual community is comparable at baseline, it will remain so throughout the study. This may not be the case; changes may occur during the intervention period that make a community more or less comparable to the intervention community.³ This limitation is shared with any method for identifying a comparison at baseline. As noted above, this became particularly salient to our example given the Flint water crisis and resulting resource allocation. Future research is needed to adapt these methods for changes across time. Moreover, in our data sources, not all features were measured during the optimal baseline year, necessitating the use of proxy years.

However, the intervention to be evaluated was complex and involved many changes over the years, and even the post-baseline proxy years were from before many of the changes were implemented.

Finally, the use of K -means clustering in general comes with its own set of assumptions and limitations.²⁰ One issue of the K -means clustering method is that the resulting assignments depend on the random starting point. The K -means algorithm gives local minima and does not guarantee to give the global minimum, so the starting points should be varied to examine the end partitioning. Another issue of the K -means algorithm is that a variable with high variability can dominate the cluster analysis. A common solution is to standardize variables, but standardizing could also hide the true groupings in the data. This is a case-by-case decision depending on the type of data and the nature of the groups. The number of clusters K is the tuning parameter that can be chosen by cross-validation if there are large number of independent and identically distributed observations, which was not possible in our presented example.

Practice Implications

K -means clustering provides a rigorous, data-driven approach to selecting one or more counterfactual groups for a community-level intervention, representing an improvement over subjective selection of comparison communities. Although it provided a consistent comparison county solution, it did not result in an optimal comparison for the intervention county. The method would be potentially more useful at other geographic levels or for other counties in or outside of Michigan. Nonrandomly assigned community interventions are common in population health; the present analysis can stimulate conversation about how best to select appropriate comparisons. In particular, it demonstrates that selection of counterfactual communities should be objective, transparent, and examined critically using a health equity lens.

Acknowledgements

The authors thank the Michigan Department of Health and Human Services for access to their Health Services Data Warehouse and consultation from the Maternal and Infant Health Division and the Division for Vital Records and Health Statistics.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This project was supported by the Agency for Healthcare Research and Quality [grant number R18-HS020208]; and the Spectrum Health-Michigan State University Alliance Corporation. The content is solely the responsibility of the authors and does not necessarily represent the official views of either funding organization.

References

1. Oakes JM. The (mis)estimation of neighborhood effects: causal inference for a practicable social epidemiology. *Soc Sci Med*. 2004;58(10):1929-1952.
2. van der Laan MJ, Peterson M, Zheng W. Estimating the effect of a community-based intervention with two communities. *J Causal Inference*. 2013;1(1):83-106.
3. Farrell AD, Henry D, Bradshaw C, Reisch T. Designs for evaluating the community-level impact of comprehensive prevention programs: examples from the CDC Centers of Excellence in Youth Violence Prevention. *J Primary Prevent*. 2016;37:165-188.
4. Sanson-Fisher RW, Bonevski B, Green LW, D'Este C. Limitations of the randomized controlled trial in evaluating population-based health interventions. *Am J Prev Med*. 2007;33(2):155-161.
5. Cook T, Shadish W, Wong V. Three conditions under which experiments and observational studies produce comparable estimates: new findings from within-study comparisons. *J Policy Analysis and Management*. 2008;27(4):724-750.
6. Bellinger C, Mohamed Jabbar MS, Zaiane O, Osornio-Vargas A. A systematic review of data mining and machine learning for air pollution epidemiology. *BMC Public Health*. 2017;17(1):907.
7. Yan S, Kwan YH, Tan CS, Thumboo J, Low LL. A systematic review of the clinical application of data-driven population segmentation analysis. *BMC Med Res Methodol*. 2018;18(1):121.
8. Wallace M, Sharfstein JM, Kaminsky J, Lessler J. Comparison of US county-level public health performance rankings with county cluster and national rankings. *JAMA Netw Open*. 2019;2(1):e186816.
9. Centers for Disease Control and Prevention, United States Department of Health and Human Services. Peer county methodology used by the Community Health Status Indicators 2015 web application. Available from: URL: <https://www.countyhealthrankings.org/sites/default/files/resources/CHSIpeerMethodology.pdf>. [cited 2020 Apr 30].
10. United States Census Bureau. American Community Survey (ACS). Available from: URL: <https://www.census.gov/programs-surveys/acs/> [cited 2020 Apr 30].
11. Health Resources and Services Administration, United States Department of Health and Human Services. Area Health Resources Files. Available from: URL: <https://data.hrsa.gov/topics/health-workforce/ahrf> [cited 2020 Apr 30].
12. University of Wisconsin Population Health Institute. County Health Rankings & Roadmaps. Available from: URL: <http://www.countyhealthrankings.org> [cited 2020 Apr 30].

13. Rettenmaier AJ, Wang J. What determines health: a causal analysis using county level data. *Eur J Health Econ*. 2013;14:821-834.
14. Roman L, Raffo JE, Zhu Q, Meghea CI. A statewide Medicaid enhanced prenatal care program: impact on birth outcomes. *JAMA Pediatr*. 2014;168(3):220-227.
15. Meghea CI, You Z, Raffo J, Leach RE, Roman LA. Statewide Medicaid enhanced prenatal care programs and infant mortality. *Pediatrics*. 2015;136(2):334-342.
16. Meghea CI, Raffo JE, VanderMeulen P, Roman LA. Moving toward evidence-based federal Healthy Start program evaluations: accounting for bias in birth outcomes studies. *Am J Public Health*. 2014; 104 Suppl 1:S25-27.
17. Krieger N. A glossary for social epidemiology. *J Epidemiol Community Health*. 2001;55:693-700.
18. McLachlan GJ. Cluster analysis and related techniques in medical research. *Stat Methods Med Res*. 1992;1(1):27-48.
19. Makles A. How to get the optimal K -means cluster solution. *The Stata Journal*. 2012;12(2):347-351.
20. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd edition. New York: Springer; 2009.
21. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366:447-453.
22. Chouldechova A. Fair prediction with disparate impact: a study of bias in recidivism prediction instruments. 2017. Available from: URL: <https://arxiv.org/abs/1703.00056> [cited 2020 Apr 30].
23. Chouldechova A, G'Sell M. Fairer and more accurate, but for whom? 2017. Available from: URL: <https://arxiv.org/abs/1707.00046> [cited 2020 Apr 30].

Table 1. County-level features used to characterize Michigan counties from the American Community Survey, Area Health Resources Files, Michigan Medicaid Vital Records, and University of Wisconsin Population Health Institute County Health Rankings, 2008-2011

<i>Domain</i>	<i>Characteristic</i>	<i>Feature</i>	<i>Source</i>	<i>Year</i>	<i>Mean</i>	<i>SD</i>
Population Composition	Race/ethnicity	% Black females age 15-54	AHRF	2010	2.20	4.54
		% Latina females age 15-54	AHRF	2010	1.93	1.65
	Urbanity	% urban	AHRF	2010	38.36	27.40
Social Economic	Social support	% families with female head	AHRF	2010	15.14	3.76
		% adults without social/emotional support ^b	UWPHI	2010	17.19	3.72
	Education	% 25+ with < high school 06-10 average	AHRF	2010	12.51	3.27
		% Black 25+ with < high school diploma	AHRF	2009	22.27	17.22
		% Latinx 25+ with < high school diploma	AHRF	2009	26.84	15.90
	Income/Poverty	Median household income (log transformed)	AHRF	2009	10.62	0.17
		% deep poverty (<50% FPL)	AHRF	2009	7.05	2.41
		% children <18 in deep poverty	AHRF	2009	10.52	4.04
		# female householder under poverty (5YR)	ACS	2009	1803.71	5692.37
	Employment	% 16-64 unemployed, 06-10 average	AHRF	2010	25.86	6.07
Access to care	Insurance	% female 18-64 uninsured	AHRF	2009	15.92	2.23
		# Medicaid-eligible female	AHRF	2008 ^a	13968.12	37053.01
		# Medicaid-eligible age<21	AHRF	2008 ^a	13871.42	36826.43
	Healthcare providers	# non-federal MD,DO in patient care /1K population	AHRF	2010	1.52	1.36
		# hospitals	AHRF	2010	2.12	3.42
		# hospital beds for obstetric care	AHRF	2011 ^a	20.71	53.45
		# FQHC / 1000 persons	AHRF	2009	1.60	2.92
Primary care provider / 100K population	UWPHI	2010	81.97	50.97		
Health outcome	Mortality	Infant mortality (total) 5-yr average 06-09 / 1000 persons	AHRF	2009	6.83	2.44
		Infant mortality (nonwhite) 5-year average / 1000 persons ^b	AHRF	2009	11.20	5.55
	Births	# Medicaid births	MMVR	2009	840.04	2121.30
		# black Medicaid births	MMVR	2009	224.48	1119.43
		% Medicaid births to unmarried women	MMVR	2009	56.43	7.22
		% Medicaid births to mothers <18	MMVR	2009	3.92	2.00
		% Black Medicaid preterm birth	MMVR	2009	6.49	14.62
		% Black Medicaid low birth weight	MMVR	2009	5.03	9.54
Health behavior	Tobacco use	% adult current smokers ^b	UWPHI	2010	23.58	4.31
	Alcohol	% binge or heavy drinking ^b	UWPHI	2010	17.65	3.66

	Sexual behavior	Chlamydia rate / 100K population	UWPHI	2010	187.55	158.49
Physical environment	Housing	Ratio of vacant vs occupied housing units	AHRF	2009	0.43	0.37
		% zip codes without healthy food outlets	UWPHI	2010	51.70	18.91
	Air quality	Air pollution days	UWPHI	2010	3.63	4.15

^a Neither 2009 nor 2010 data are available.

^b These variables have missing values for certain counties (n missing for % adults without social/emotional support = 18, infant mortality (nonwhite) 5-year average = 35, % adult current smokers = 18, % binge or heavy drinking = 14), which were replaced by the mean values.

ACS = American Community Survey

AHRF = Area Health Resources Files

DO = Doctor of Osteopathy

FPL = Federal Poverty Level

FQHC = Federally Qualified Health Center

MD = Medical Doctor

MMVR = Michigan Medicaid Vital Records, i.e., a limited dataset of Michigan resident live birth records for Medicaid-insured births only, retrieved from the Michigan Department of Health and Human Services (MDHHS) Health Services Data Warehouse.

SD = standard deviation

UWPHI = University of Wisconsin Population Health Institute County Health Rankings

Table 2. Number of times a county is grouped together with Kent County in 1,000 iterations, Michigan, 2008-2011

County	<i>K=3</i>	<i>K=4</i>	<i>K=5</i>	<i>K=6</i>
Genesee	1,000	982	239	127
Macomb	1,000	1,000	1,000	1,000
Oakland	1,000	986	985	979

Table 3. Features for Kent, Macomb, Oakland, and Genesee Counties, Michigan, 2008-2011

Feature	Kent	Macomb	Oakland	Genesee
% Black females age 15-54	7.3	7.0	10.5	14.9
% Latina females age 15-54	6.7	1.5	2.4	2.0
% urban	84.3	97.2	95.2	83.2
% families with female head	18.8	18.9	16.9	26.0
% adults without social/emotional support	18.2	20.2	17.3	23.0
% 25+ with < high school 06-10 average	11.7	12.4	7.8	11.9
% Black 25+ with < high school diploma	19.7	13.5	10.3	16.3
% Latinx 25+ with < high school diploma	46.3	22.4	27.0	16.9
Median household income (log transformed)	10.8	10.8	11.0	10.6
% deep poverty (<50% FPL)	6.8	4.8	4.3	9.5
% children <18 in deep poverty	9.6	7.0	5.7	15.5
# female householder under poverty (5YR)	8864.0	7333.0	9024.0	9745.0
% 16-64 unemployed, 06-10 average	19.5	22.7	20.8	30.7
% female 18-64 uninsured	14.8	14.7	12.3	13.0
# Medicaid-eligible female	72009.0	74763.0	79324.0	65315.0
# Medicaid-eligible age<21	74313.0	73401.0	78119.0	65053.0
# non-federal MD,DO in patient care /1K population	3.3	1.6	5.7	2.5
# hospitals	7.0	8.0	17.0	4.0
# hospital beds for obstetric care	177.0	55.0	255.0	90.0
# FQHC / 1000 persons	11.0	1.0	2.0	5.0
Primary care provider / 100K population	129.3	71.0	207.5	132.1
Infant mortality (total) 5-yr average 06-09 /1K population	7.4	6.5	6.4	9.2
Infant mortality (nonwhite) 5-year average /1K population	13.7	10.7	10.8	16.9
# Medicaid births	5045.0	4830.0	5320.0	3739.0
# black Medicaid births	1005.0	947.0	1506.0	1407.0
% Medicaid births to unmarried women	60.4	52.8	57.3	71.8
% Medicaid births to mothers <18	4.7	2.9	3.5	6.3
% Black Medicaid preterm birth	13.5	12.8	14.3	17.0
% Black Medicaid low birth weight	12.1	13.2	12.7	15.2
% adult current smokers	21.3	22.5	18.4	25.9
% binge or heavy drinking	17.1	18.6	16.0	16.0
Chlamydia rate / 100K population	559.9	195.8	274.7	709.3
Ratio of vacant vs occupied housing units	0.1	0.1	0.1	0.1
% zip codes without healthy food outlets	55.1	14.6	33.3	45.9
Air pollution days	15.0	9.0	10.0	8.0

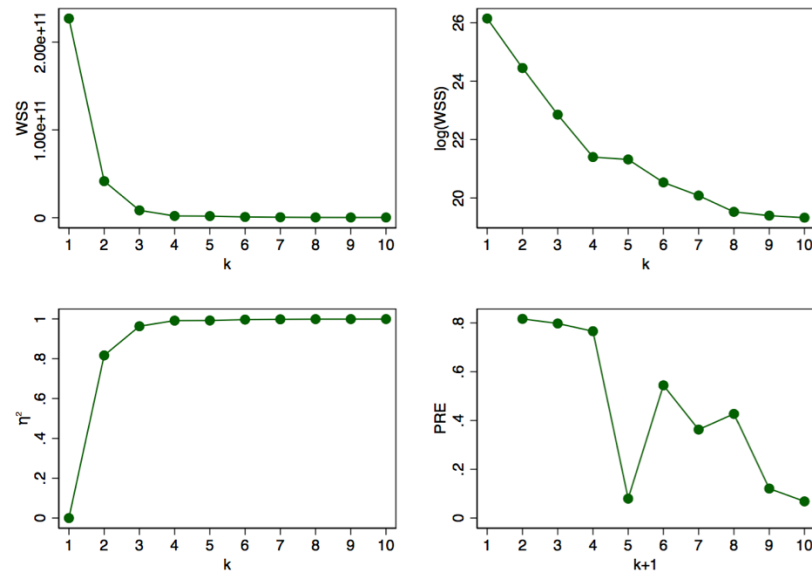
DO = Doctor of Osteopathy

FPL = Federal Poverty Level

FQHC = Federally Qualified Health Center

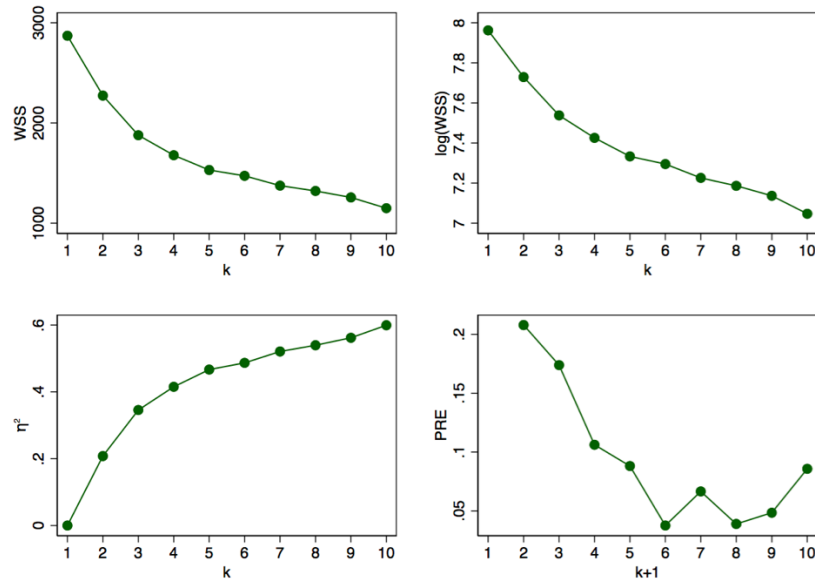
MD = Medical Doctor

Figure 1. WSS, $\log(\text{WSS})$, η^2 coefficient and PRE for 10 cluster solutions when features are not standardized, Michigan, 2008-2011



PRE = proportional reduction of error
WSS = within sum of squares

Figure 2. WSS, log(WSS), η^2 coefficient and PRE for 10 cluster solutions when features are standardized, Michigan, 2008-2011



PRE = proportional reduction of error
WSS = within sum of squares