Assessing U.S. County-level Mental Health Outcomes: A Spatial Analysis of Quality of Life

Marina Armendariz

Department of Biobehavioral Health

The Pennsylvania State University

Stephen A. Matthews

Departments of Sociology and Anthropology

The Pennsylvania State University

Note: We are currently preparing this manuscript for submission to a journal with the goal to complete this task by December 2018. We are currently checking the sensitivity of results from the spatial econometric models and all maps will be recreated in ArcGIS for a PAA poster. Sections of the paper, including the discussion and conclusions, are currently being expanded.

Abstract

Social and economic conditions are often cited as fundamental determinants of various health outcomes and evidence suggests that they also show spatial clustering. While there is extensive research on the spatial patterning of life expectancy, mortality and chronic health outcomes; there is limited work examining the spatial distribution of subjective health, such as quality of life. The current study explores the spatial patterning for three county-level self-reported health (SRH) outcomes: poor/fair health (rates), physical distress days (frequency), and mental distress days (frequency) as well as social and economic conditions, including average county household income and county unemployment rates. Various spatial tools and methods were utilized, including a combination of exploratory spatial data analysis (ESDA) and spatial econometric techniques. Findings from the spatially informed analysis revealed spatial autocorrelation, and thereby spillover effects, in key explanatory and outcome variables. Moreover, findings from analytic models support the notion that county-level disadvantaged socioeconomic status contributes to health inequalities at the population level. Lastly, spatial regimes revealed that there were regional differences in quality of life health outcomes between metro and non-metro counties. To begin to improve population health, the spatial patterning and neighboring effects of health outcomes along with the associated risk factors should be considered at various levels of geography.

Keywords: quality of life; mental health; spatial analysis; spatial econometrics; socioeconomic status

Introduction

At the individual level, there is a strong association between socioeconomic status and health, particularly socioeconomic disadvantage and poor health. This persistent finding is also observed at the county-level. For example, poverty rates, occupation, wage and overall a lack of economic resources contribute to health inequalities experienced, especially in rural counties (Probst et al., 2004; Krieger et al., 2005). A rural health disadvantage has been established in the United States. Rural counties have been known to be disadvantaged along a series of health outcomes, including premature mortality (before age 75), unintentional injuries, suicide, and some chronic diseases (Eberhardt & Pamuk, 2004). More recent research is finding that such a health disadvantage continues to exist and also exists for subjective measures of health, such as self-reported health (Dwyer-Lindgren et al., 2017).

However, when examining county-level data, implications of the findings should be interpreted with caution because they can be overestimated due to the existence of spatial heterogeneity. For instance, studies examining spatial clusters of county-level chronic conditions, such as diabetes and obesity, concluded that areas along the Southern belt and neighboring counties were highly concentrated and spatially associated (Myers et al., 2017; Shrestha et al., 2012). Additionally, social and economic conditions are often cited as fundamental determinants of various health outcomes and evidence suggests that they also show spatial clustering. Krieger and colleagues (2005) used geographic and poverty measures to help "paint the picture" of US socioeconomic inequalities in health.

While there is extensive research on the spatial patterning of life expectancy, mortality and chronic health outcomes (Tabb et al., 2018); there is limited work examining the spatial distribution of subjective health, such as quality of life. A recent study by Dwyer-Lindgren and colleagues (2017) used county-level data to explore spatial patterns of self-reported physical and mental health where they found pronounced disparities in rural areas. In turn, this paper aims to expand on this literature by exploring the spatial patterning for three county-level self-reported health (SRH) outcomes: poor/fair health (rates), physical distress days (frequency), and mental distress days (frequency) as well as social and economic conditions, including average county household income and county unemployment rates.

Moreover, various risk factors contribute to the compounding effects of rurality on health. For instance, rural racial/ethnic minorities are considered to be disadvantaged in health compared to other rural racial ethnic minorities as well as urban minority groups (Probst et al., 2004). As a result, the present study will account for demographic covariates including race/ethnicity and the female population. Nonetheless, socioeconomic disadvantage may be a larger and arguably more important characterization of the compounding effects of rurality on health.

Therefore, the present study aims to expand on the work of quality of life health outcomes while considering the contribution of social/economic condition and demographic covariates. Assessing counties as the unit of analysis is a practical way to encompass the entire U.S., including metropolitan and non-metropolitan areas. Thus, the following research questions are examined using a spatial approach:

1) Are social/economic conditions as well as measures of quality of life spatially clustered across US counties?
2) What is the association between social/economic conditions and quality of life health outcomes?
    a. Do regional variations exist, particularly between metro and non-metro counties?

Methods

Measures

Data Sources

The present paper used the 2016 County Health Rankings (CHR) data. The CHR data are used to assess the overall health of each county across the United States. Annually since 2010, the Rankings has compiled county-level measures from national and state data sources to identify both health factors and health outcomes that contribute to the nation's health (Remington et al., 2015), resulting in aggregated county-level estimates. The Economic Research Service 2013 Rural-Continuum Codes (RUCC) to distinguish between metropolitan and non-metropolitan counties.

Health Outcome Variables

<u>SRH: Poor/Fair Health, Mental Distress and Physical Distress</u> The original data source of these measures are from the 2015 Behavioral Risk Factor Surveillance System (BRFSS). For poor/fair health, the CHR provides an aggregate value for each county resulting in the percentage of adults reporting fair or poor health. For the purpose of the present study, this variable was left as an aggregate value (0-100%). To determine frequency of mental and physical distress, responses are based on self-report of how many days in the past 30 days have respondents experienced poor physical or mental distress. The CHR uses the average number of days of a county's adult respondents report having poor physical and/or mental health.

Key Explanatory Variables

<u>Socioeconomic Status</u> Median household income was used as one measure of socioeconomic status. The original data source of this measure is from the 2015 Small Area Income and Poverty Estimates. The median household income is provided by the CHR based on a county's income

where half of the households in the county earn more and half of the county households earn less. Non-logged median household income is used to report descriptive county characteristics and for spatial analysis. Unemployment was also used to assess socioeconomic status. It refers to the percentage of the civilian labor force of individuals 16 and older who are unemployed but seeking employment. Percent unemployed ranges from 0 to 100%.

Metro/Non-Metro Using the Rural-Urban Continuum Codes (RUCC), counties were dummy coded as metropolitan or non-metropolitan, metro and non-metro hereafter. Each U.S. county is assigned one of nine codes. The official Office of Management and Budget (OMB) metro and non-metro have been subdivided into three metro (1-3) and six non-metro (4-9) categories (USDA, 2016). Thus, counties with a RUCC ranging from 1-3 was dummy coded as 0 and counties ranging from 4-9 were dummy coded as 1.

Analysis

This paper aimed to understand the relationship between county-level socioeconomic status (1) average household income and 2) unemployment rate) and three county-level self-reported measures of health (1) percent fair/poor health; 2) frequency of physical distress and 3) frequency of mental distress). The methodology used in this paper applies a combination of exploratory spatial data analysis (ESDA) and spatial econometric techniques. All analyses were conducted using the *GeoDa* and *GeoDa Space* software (Anselin, Syabri, & Kho, 2006).

To address research question one, a serious of ESDA techniques were conducted to identify the spatial clustering of the key independent and dependent variables. ESDA is a useful technique for exploratory purposes and in searching for spatial regimes (Baller et al., 2001). Specifically, ESDA was applied to 1) visually display outcomes of interest on a map, 2) create spatial weights matrices for Moran's I statistics and 3) identify significant High-High (H-H) and Low-Low (L-L) clustering using LISA statistics. Moran's I and LISA reports were useful for identifying levels of spatial autocorrelation globally and locally, respectively.

Results from ESDA were important in informing spatial econometric techniques, including standard OLS regressions and Spatial Lag regressions. To address research question two, three OLS regressions were performed using the spatial weights that were previously created (Queen 1) for all social/economic variables and demographic variables against each health outcome. Results from the OLS models were compared to results from Spatial Lag regression models in order to find the best fit model. This step is necessary when ESDA and diagnostic regression analyses confirm the existence of spatial dependence/heterogeneity (Baller et al., 2001; Tabb et al., 2018).

Lastly, spatial regimes were employed to determine whether there were regional differences between metro and non-metro areas. The application of spatial regimes has the potential to elucidate different social mechanisms by region or different relative significance of the covariates in the model (Baller et al., 2001).

Results

Descriptive Statistics

Table 1 provides the descriptive statistics generated for 3,109 U.S. counties. It should be noted that all counties in Alaska and Hawaii were removed due to incomplete data. The data for the three dependent variables across all U.S. counties indicates that the mean percentage of adults in the U.S. reporting fair to poor health is 16.92% (SD 4.97). The mean number of days of physical distress and mental distress was 11.5 (2.1) and 11.2 (2.5), respectively. The median household income is $47,117. The mean unemployment rate was 6.02 (2.3). The average percentage of Hispanics is 9.02%, Blacks is 8.02 % and female is 49.9%. Hispanic and Black county-level percentages are not normally distributed. Further, bivariate scatter plots were also conducted (not shown) and both socioeconomic measures indicated a significant association with the three self-rated health outcome variables; however, there is some evidence of a curvilinear relationship between household income and all three outcome variables. This issue will be adjusted in a subsequent version of this study.

Table 1. Descriptive Statistics for all U.S. Counties

	All Counties		
N	3,109	Min.	Max
Dependent Variables			
Health Outcomes (M, SD)			
Fair to poor health	16.92 (4.97)	0	41.7
Physical Distress (days)	11.5 (2.5)	0	21.7
Mental Distress (days)	11.2 (2.1)	0	19.2
Explanatory Variables			
Socioeconomic Status (M, SD)			
Median household income ($000)	47,117 (12,103)	$21,658	$125,635
% Unemployment	6.02 (2.3)	0	23.61
Demographic Covariates (%)			
% Hispanic	9.02	0%	95%
% Black	8.02	0%	85%
% Female	49.9	0%	56%

Visualization

Figures 1 and 2 are visual illustrations of the spatial distribution and spatial clustering of the key explanatory and outcome variables. In Fig. 1, there is some indication of spatial concentration of poor/fair health, physical distress, and mental distress. For example, higher rates and frequencies, as indicated by the darker areas, are seen along the Southern belt, while less is seen in parts of the Midwest. Additionally, for unemployment, there are patterns of higher rates along the coastal areas of the West and South. However, this socioeconomic condition does not map comparably to household income where there is less evident patterning on the West. In other words, spatial concentrations for household income are less pronounced, but they are present. For example, lowest incomes are seen along neighboring areas of the mid-inland areas of the U.S. Metropolitan counties may account for this, which warrants future exploration.

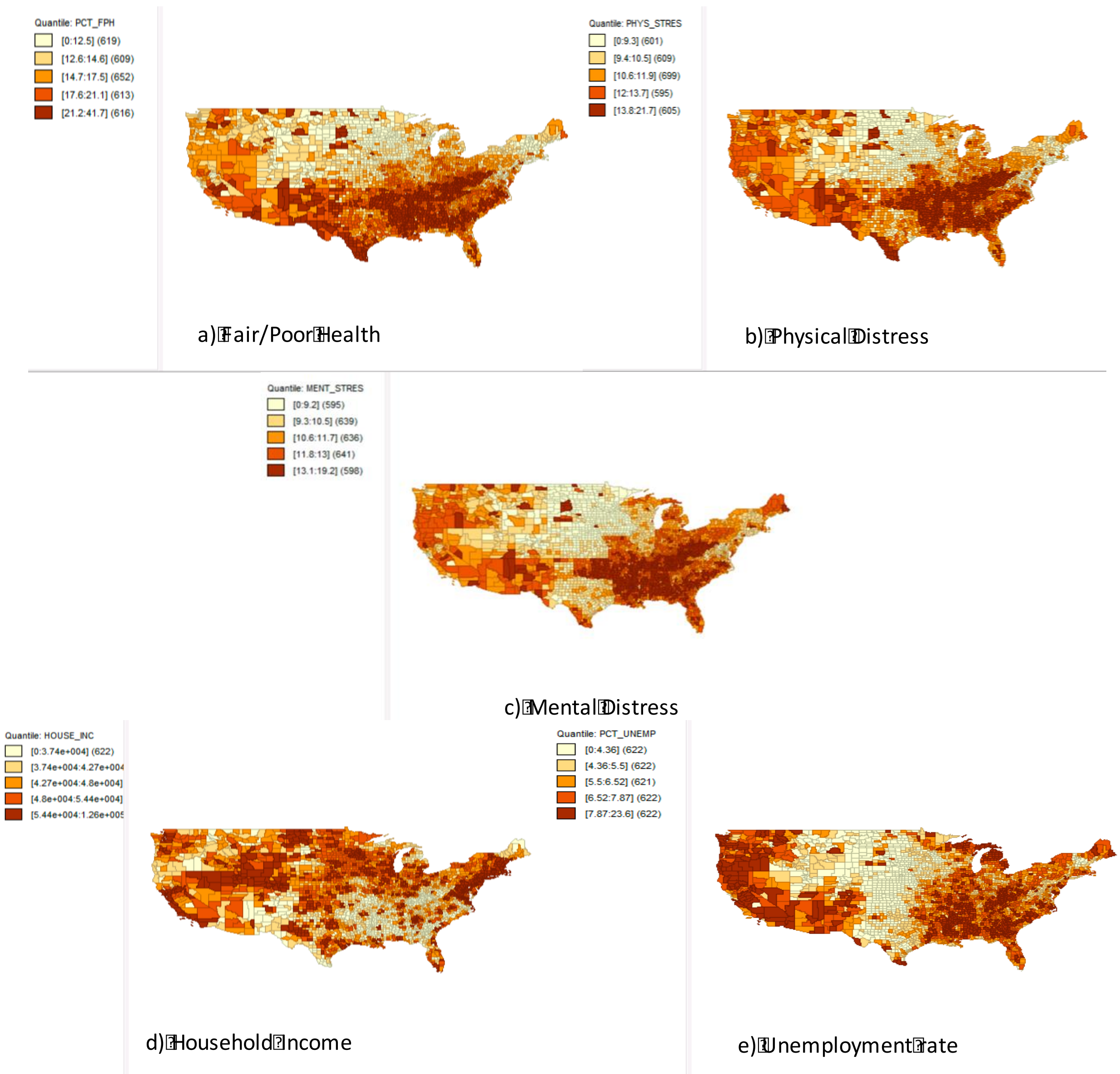

Figure 1. Mapped visualization of key outcome variables: a) poor/fair health, b) physical distress, and c) mental distress and key variables of interest: d) household income and e) unemployment

Global Moran's I

The global spatial association was examined using Moran's *I* statistics for spatial autocorrelation. The results for all variables of interest are given in Table 2. Psuedo p-values were generated in the Moran's scatter plot, where permutations of 999 were tested. All tests were statistically significant, $p<.001$. The results suggest that measures of self-reported health and socioeconomic status have an organized spatial pattern. Thus, spatial structuring should be taken into account, particularly when performing spatial regression analyses.

Table 2. Moran's *I* statistic for spatial autocorrelation in SRH and SES

Spatial Weights	*Poor/Fair Health*	*Physical Distress*	*Mental Distress*	*Unemployment*	*Household Income*
Queen	.703***	.675***	.726***	.649***	.587***
Rook	.687***	.678***	.729***	.649***	.591***

***: $p< .001$

Local LISA mapping

Additionally, the local spatial autocorrelation was examined using the local Moran's statistics. LISA mapping is a visual indicator for revealing the extent to which the pattern of a value at that location is and the values in the neighboring locations are compatible with spatial randomness (Baller et al., 2001). Figure 2., indicates local clustering of high clustering surrounded by high (High-High/H-H) and of low clustering surrounded by low (Low-Low/L-L). These areas were significant at the $p<.01$ and $p<.05$ levels, thus, rejecting the null hypothesis that there is no clustering.

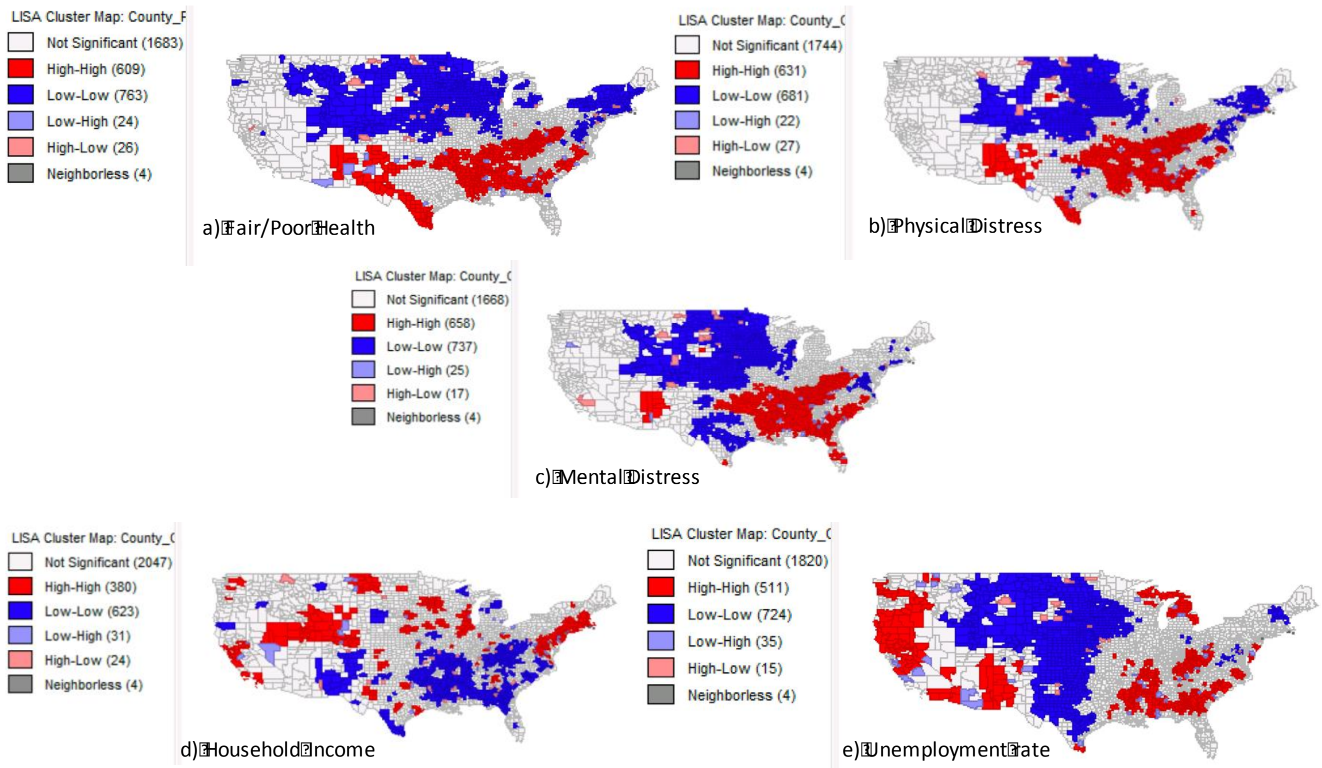

Figure 2. Mapped visualization of LISA cluster maps (Queen weights). Key outcome variables: a) poor/fair health, b) physical distress, and c) mental distress and key variables of interest: d) household income and e) unemployment

Spatial econometrics

Spatial regression models were conducted to account for the presence of spatial autocorrelation. Using a weights matrix specification (Queen1), OLS regressions were performed independently for each dependent variable. Further, spatial diagnostics from the OLS models revealed issues of multicollinearity as well as significant Lagrange multipliers (lag and errors), suggesting support for the use of Spatial lag models. Additionally, Figure 3 shows residual

Moran's *I* plots from spatial lag models, which indicate that spatial autocorrelation was reduced.

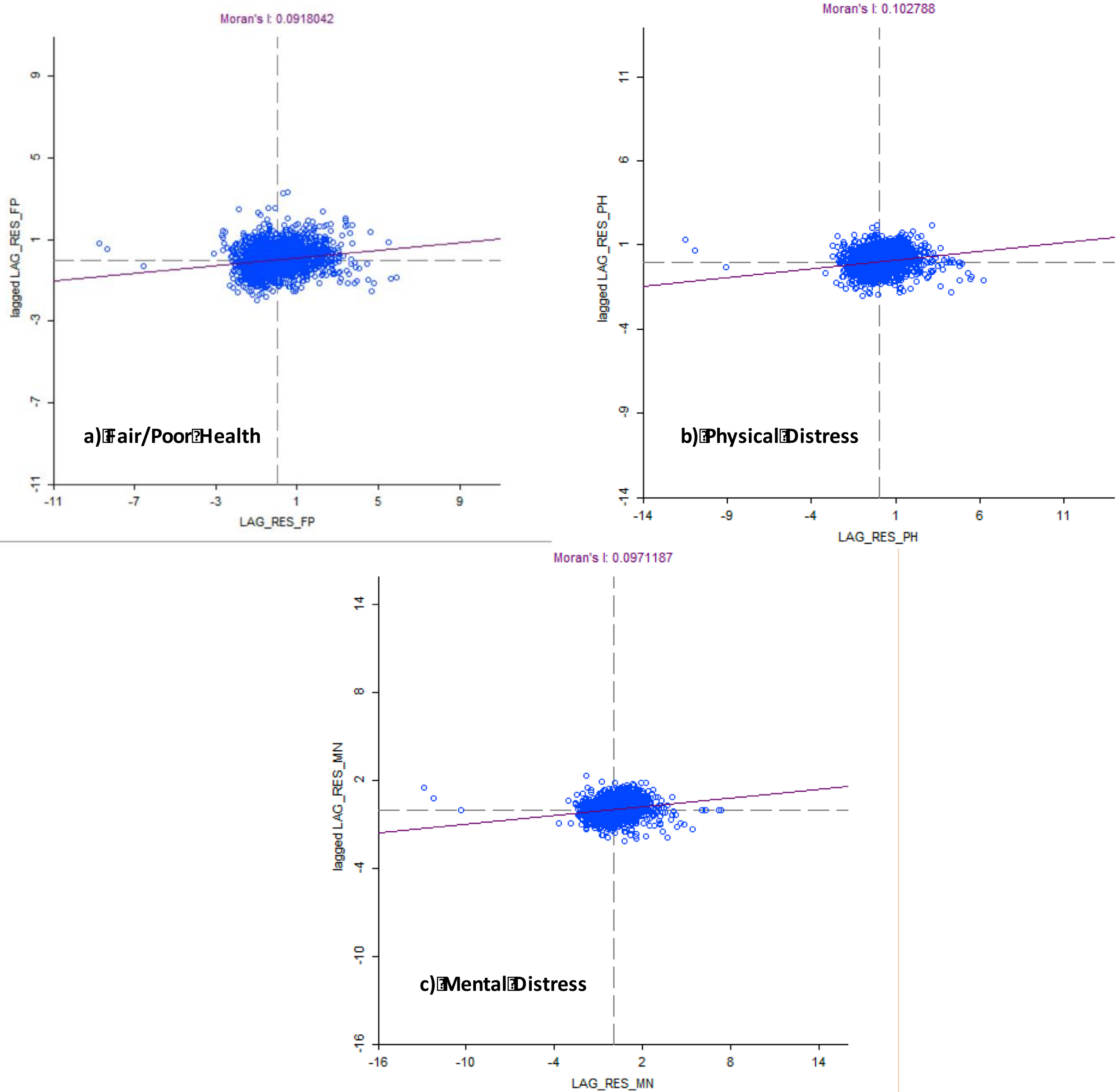

Figure 3. Residual plots from spatial lag models

Tables 3-5 address research question two. Table 3 compares regression results, fit statistics, and spatial diagnostics for poor/fair health. The OLS regression indicates that the overall model is statistically significant, (F = (6, 3103) = 689.5, Adj-R2 = .58, p < .001). Household income is significantly and negatively associated with fair/poor health; counties with higher average income have lower rates of fair/poor health reports. Unemployment was a significant and positive predictor of fair/poor health. In the spatial lag model, results were similar; however, there was a significant decrease in the beta coefficient for household income. Additionally, the log likelihood and AIC improved, suggesting that the spatial lag is a better fit model.

Table 3. Fully Adjusted Regression for U.S. Counties for Fair/Poor Health

	OLS		Spatial Lag	
	Estimate	SE	Estimate	SE
Explanatory Variables				
Constant	4.66*	.441	3.41*	.782
Socioeconomic Status				
Median household income ($000)	-9.60*	8.36	-.0001*	4.18
Unemployment rate	.972*	.030	.321*	.022
Demographic Controls				
% Hispanic	.131*	.004	.086*	.003
% Black	.121*	.004	.059*	.003
% Female	.091*	.091	.157*	.015
Fit Statistics				
AIC	16,510.2		13,936	
Log likelihood	-8,249.1		-6,960.9	
Spatial Diagnostics				
Moran's I	.640*			
Lagrange Multiplier (lag)	2,414*			
Lagrange Multiplier (error)	3,671*			

* = Significant at $p<.05$

Table 4 compares regression results, fit statistics, and spatial diagnostics for physical distress. The OLS regression indicates that the overall model is statistically significant (F (6, 3103) = 1,103, Adj-R2 = .63, p <. 001). The log likelihood and AIC in the spatial lag regression indicates that the overall model was improved. Although the beta coefficient decreased from the OLS model to the spatial lag model, there was a positive association between unemployment rates and average physical days of distress; frequency of physical distress increased with unemployment percent increases (p<.001). Household income also remained statistically and negatively significant in the spatial lag model; less household income was associated with more days of physical distress. In the spatial lag, the beta coefficient for household income significantly increased.

Table 4. Fully Adjusted Regression for U.S. Counties for Physical Distress

	OLS		Spatial Lag	
	Estimate	SE	Estimate	SE
Explanatory Variables				
Constant	6.85*	.506	6.40*	.014
Socioeconomic Status				
Median household income ($000)	-.0001*	2.55	-6.75*	2.34
Unemployment rate	.370*	.014	.219*	.012
Demographic Controls (%)				
% Hispanic	.029*	.002	.023*	.001
% Black	.022*	.002	.012*	.001
% Female	.134*	.010	.117*	.007
Fit Statistics				
AIC	11,299.1		9,936	
Log likelihood	-5,643.5		-4,961.4	
Spatial Diagnostics				
Moran's I	.55*			
Lagrange Multiplier (lag)	1,718*			
Lagrange Multiplier (error)	3,671*			

* = Significant at $p<.05$

Table 5 compares regression results, fit statistics, and spatial diagnostics for mental distress. The OLS regression indicates that the overall model is statistically significant (F (6, 3103) = 11,009, Adj-R2 = .61, p <. 001). The improved log likelihood and AIC in the spatial lag regression indicates that the overall model was improved. The association between the explanatory variables and mental distress are similar to those of physical distress. There was a positive association between unemployment and frequency of mental distress and a negative association between household income and frequency of mental distress. All were significant at the .01 level.

Table 5. Fully Adjusted Regression for U.S. Counties for Mental Distress

	OLS		**Spatial Lag**	
	Estimate	SE	Estimate	SE
Explanatory Variables				
Constant	4.66*	.441	-1.04*	.339
Socioeconomic Status				
Median household income ($000)	-7.03*	2.22	-4.47*	1.74
% Unemployment	.383*	.012	.193*	.009
Demographic Controls				
% Hispanic	-.001	.002	.005*	.001
% Black	.017*	.002	.008*	.001
% Female	.146*	.008	.129*	.006
Fit Statistics				
AIC	10,442.8		8,602.5	
Log likelihood	-5,215.4		-4,294.2	
Spatial Diagnostics				
Moran's I	.55*			
Lagrange Multiplier (lag)	2,153*			
Lagrange Multiplier (error)	3,077*			

* = Significant at *p<.05*

Spatial Regimes

ESDA and regression results supported the search for spatial regimes. The spatial regimes allow the coefficients to be different in each regime (Baller et al., 2001), which have been divided by metro/non-metro status in the present study. Table 6 displays the basic OLS regressions results for each regime. The Chow test is a test of coefficient differences (Baller et al., 2001). Specifically, the overall or global test indicates that there are regional differences by metro non-metro status. In general, the individual coefficients of each explanatory variable and covariates indicate that there are significant differences in the coefficients, except for unemployment. For unemployment, there was a marginal coefficient difference between metro and non-metro.

Table 6. Regression Coefficients using Spatial Regimes—Metro/Non-Metro Status

		Metro		Non-Metro	
	Chow Test	Coefficient	SE	Coefficient	SE
Fair to Poor Health					
Global Test	**234.1***				
Constant	**55.45***	5.71*	1.21	20.15*	1.5
Socioeconomic Status					
Median household income ($000)	**99.25***	-.0001*	<.001	-.0002*	<.001
% Unemployment	1.21	.338*	.048	.405*	.035
Demographic Controls (%)					
% Hispanic	**15.03***	.103*	.006	.133*	.004
% Black	**8.40***	.082	.005	.105*	.005
% Female	**29.24***	.304	.025	.098*	.027
Physical Distress					
Global Test	**261.4***				
Constant	**104.5***	3.11*	.627	13.44*	.792
Socioeconomic Status					
Median household income ($000)	**114.0***	-<.001	<.001	-<.001	<.001
% Unemployment	2.97+	.269*	.003	.323*	.018
Demographic Controls (%)					
% Hispanic	3.53	.022*	.003	.030*	.002
% Black	.221	.016*	.003	.018*	.002
% Female	**67.02***	.210*	.013	.050*	.014
Mental Distress					
Global Test	**216.8***				
Constant	**87.57***	1.83*	.551	10.13*	.694
Socioeconomic Status					
Median household income ($000)	**58.62***	-<.001*	<.001	-<.001*	<.001
% Unemployment	3.05+	.302*	.022	.350*	.016
Demographic Controls (%)					
% Hispanic	**4.31***	-.007*	.002	-<.001	.002
% Black	.918	.010*	.002	.014*	.002
% Female	**70.81***	.216	.011	.071	.012

* = Significant at $p<.05$

Discussion

This paper aimed to display the spatial patterning of county-level socioeconomic status and self-reported health. Illustrative results demonstrate that there is some indication of spatial patterning of three self-reported health outcomes: poor/fair health, frequency of physical distress (days), and frequency of mental distress (days). This is supported by recent research that has aimed to elucidate geographic variation of county-level self-reported health (Dwyer-Lindgren et al., 2017). Moreover, this study also explored spatial patterns of socioeconomic status, including unemployment rates and average household incomes. The spatial patterns are less visible visually; however, Moran's *I* confirmed that key explanatory and outcome variables are spatially auto correlated. These findings support the importance of considering spatial spillover effects in population health (Tabb et al., 2018).

Moreover, county-level socioeconomic status was associated with all county-level quality of life health outcomes. Specifically, county-level unemployment rates were positively associated with an increase in the frequency of physical and mental distress. However, higher household income average was associated with lower frequency of physical and mental distress. These findings support the notion that disadvantaged socioeconomic status contributes to health inequalities in the U.S. at the population level (Probst et al., 2004; Krieger et al., 2005). Additionally, spatial diagnostics indicated the existence of spatial autocorrelations; thus, it was determined that a spatial lag was the best fit model. However, although spatial patterning was reduced, it should be noted that spatial dependence was not completely removed. Thus, future analysis of this work merits the need for a Spatial Durbin.

Lastly, use of spatial regimes determined revealed that there were regional differences in quality of life health outcomes between metro and non-metro counties. Poorer quality of life, as measured by fair/poor health as well as frequency of physical and mental distress, was observed in more non-metro counties compared to metro counties. Previous work has found pronounced effects of poorer health in rural areas of the south. Therefore, an extension of this work warrants the need to find regional differences between the South and non-South regions (East, West, Midwest) using spatial techniques.

Limitations and Strengths

One major limitation of this study is that it only provides a cross-sectional snapshot of county-level health in 2016. For policy implications, it would be useful to include multiple years in order to examine changes in health rates. Further, the present study did not fully remedy the spatial dependence; therefore, additional analyses are required. Nonetheless, one strength of this study is the use of a macro-level lens to inform health researchers and public health officials of geographic areas that need improved social and economic conditions.

Conclusion

In order to begin to improve population health, the spatial patterning and neighboring effects of health outcomes along with the associated risk factors should be considered at various geographic levels. For example, regional, state and local levels may yield varying public health needs that would otherwise not be captured using traditional methods. Further, understanding both population health and individual-level health are necessary for targeting health inequalities across the U.S.

References

Baller, R. D., Anselin, L., Messner, S. F., Deane, G., & Hawkins, D. F. (2001). Structural covariates of US county homicide rates: Incorporating spatial effects. *Criminology*, *39*(3), 561-588.

Dwyer-Lindgren, L., Mackenbach, J. P., van Lenthe, F. J., & Mokdad, A. H. (2017). Self-reported general health, physical distress, mental distress, and activity limitation by US county, 1995-2012. *Population Health Metrics*, 15(1), 16.

Eberhardt, M. S., & Pamuk, E. R. (2004). The importance of place of residence: Examining health in rural and non-rural areas. *American Journal of Public Health*, *94*(10), 1682-1686.

Krieger, N., Chen, J. T., Waterman, P. D., Rehkopf, D. H., & Subramanian, S. V. (2005). Painting a truer picture of US socioeconomic and racial/ethnic health inequalities: The Public Health Disparities Geocoding Project. *American Journal of Public Health*, *95*(2), 312-323.

Myers, C. A., Slack, T., Broyles, S. T., Heymsfield, S. B., Church, T. S., & Martin, C. K. (2016). Diabetes prevalence is associated with different community factors in the diabetes belt versus the rest of the United States. *Obesity*.

Probst, J. C., Moore, C. G., Glover, S. H., & Samuels, M. E. (2004). Person and place: the compounding effects of race/ethnicity and rurality on health. *American Journal of Public Health*, *94*(10), 1695-1703.

Remington, P. L., Catlin, B. B., & Gennuso, K. P. (2015). The county health rankings: rationale and methods. *Population Health Metrics*, *13*(1), 11.

Shrestha, S. S., Thompson, T. J., Kirtland, K. A., Gregg, E. W., Beckles, G. L., Luman, E. T., ... & Geiss, L. S. (2016). Changes in Disparity in County-Level Diagnosed Diabetes Prevalence and Incidence in the United States, between 2004 and 2012. *PLoS One*, *11*(8), e015987

Tabb, L. P., McClure, L. A., Quick, H., Purtle, J., & Roux, A. V. D. (2018). Assessing the spatial heterogeneity in overall health across the United States using spatial regression methods: The contribution of health factors and county-level demographics. *Health & Place*, *51*, 68-77.