Beyond Geographic Connectivity: How Migration Flows Shape Overall Health Across U.S. Counties*

Noli Brazil

Community and Regional Development
Department of Human Ecology
University of California, Davis

September 12, 2018

A long line of research has demonstrated that the contextual features of a place are associated with the health and well-being of its residents (Diez Roux and Mair, 2010; Oakes et al., 2015). Juxtaposed against this body of evidence is the increased attention given to examining spatial patterns and processes that may help explain geographic health inequalities (Brazil, 2017; Sparks and Sparks, 2010; Tabb et al., 2018; Yang, et al., 2012, 2015). Much of this attention has been paid to the spillover effects of spatially proximate places; that is, places are influenced by nearby communities , or as articulated by Tobler (1970): Everything is related to everything else but near things are more related than distant things.

Although health research has drawn attention to the possibility that communities may influence, or be influenced by, physically adjacent areas, the literature to date has not explored community-to-community effects that transcend geographic proximity. The assumption underlying the spatial dependence of health and its determinants is that the influence of ecological conditions spill over or diffuse across geographic boundaries. The implication is that space matters, but that influence is constrained by

---

*Please address correspondence to: Noli Brazil, Department of Human Ecology, University of California, Davis. Email: nbrazil@ucdavis.edu.

distance and that nearest neighbors matter most. Although studies of spatial dependence have been invaluable in empirically demonstrating the importance of examining contextual effects beyond the immediate community, restricting processes of influence to operate only among contiguous neighbors ignores processes of influence that are not spatially bounded. That is, communities are connected not simply because they are geographically close to one another, but also because they interact through political, economic, social and demographic processes, which may not be spatially structured (Radil et al., 2010).

Such an effect is anticipated by a growing body of research that points to the interdependency of spatially distal communities (Bastomski et al., 2017; Hipp and Perrin, 2009; Leenders, 2002; Mears and Bhati, 2006; Tita and Radil, 2010; Wellman, 1999). For example, Tita and Radil (2011) found that violent crime diffuses across neighborhoods via pathways defined by gang rivalries. This position is also anticipated by social network theory which models interconnectedness among actors or nodes via shared membership into formal/informal groups (Browning et al., 2004). The important insight here is that communities may affect and be affected by other communities with which they coexist and interact, and these interactions are not entirely bounded by spatial proximity.

The present study analyzes one particular pathway connecting the health and well being of communities both near and far: migration flows. The justification stems in part from the observation that social conditions affecting health outcomes are not airborne phenomena that disperse across boundaries. Rather, people and groups move, taking their behaviors with them. The assumption is that individuals coming from an area with, for example, high depression levels may negatively influence the mental health of residents in their destination places both through the transmission of depression and the ecological conditions influencing depression. Furthermore, the influence of distal communities connected via migration streams is not a simple function of the individuals moving from place to place, but the underlying push and pull factors establishing these streams, such as economic, political and institutional networks (Hipp and Perrin, 2009; Tita and Radil, 2010). This position is motivated by the vast literature on social and migration networks. Social scientists have employed social network analysis in an effort to explain a number of social processes, including the diffusion of innovations, norms, and information among individuals and communities (Christakis and Fowler, 2008; Haynie, 2001; Smith and Christakis, 2008; Valente, 1996). Migration and social networks are integrally linked in that networks are a fundamental component of local social organization, and that migration produces and/or transforms a community's social networks, thereby affecting its social structure, which has consequences on health and well being in the area (Brown, 2002).

In this study, I examine the association between the health and social conditions of places connected via internal migration flows in the United States. Employing a combination of spatial regression and social network methods using data on 2011-2015 census county-to-county migration flows and health outcome and quality measures from the County Health Rankings (CHR), I evaluated the following research objectives: (1) Examine the extent to which migration flows connect near and distant counties; (2) Compare the diffusion of health outcomes across counties linked geographically and via migration networks; and (3) Compare the effects of economic and social conditions in spatially proximate and migration-connected counties on overall health.

1 Data and Methods

1.1 Data Sources

Data for this project were obtained from two primary sources: U.S. Census County-to-County Migration Flows and the Community Health Rankings (CHR), a project created by the University of Wisconsin Population Health Institute and the Robert

Wood Johnson Foundation. I used the County-to-County Migration Flows data to measure migration counts from county to county for all counties in the United States. Migration flows are estimated from the 2011-2015 American Community Survey (ACS) 5-year estimates. Migration flow counts are period estimates that measure where people lived when surveyed (current residence) and where they lived 1 year prior (residence one year ago). The data are collected continuously over a 5-year period in order to provide a large enough sample for estimates in smaller geographies. A 5-year estimate also smoothes year-to-year fluctuation in migration rates. The flow estimates resemble the annual number of movers between counties for the 5-year period data was collected. I include counties only in the contiguous United States, yielding a final analytic sample of 3,108 counties.

I used the CHR to measure the health outcome variable and economic and social mechanisms of county-to-county influence. The CHR dataset contains a wide set of health outcomes and determinants for every county in the United States. The data are compiled from over 20 national and state data sources including the American Community Survey (ACS) 5-year estimates, the Centers for Disease Control and Prevention (CDC), and the Behavioral Risk Factor Surveillance System (BRFSS). Further details about CHR methods and measures are provided elsewhere (Remington and Booske, 2011).

1.2 Dependent and Independent Variables

The dependent variable is a composite of three variables: the percent of adults reporting fair or poor health, the average number of physically unhealthy days per month, and the average number of reported mentally unhealthy days per month. These variables were taken from the 2018 CHR and measures reflect responses from the 2016 BRFSS. I used the first factor of a Principal Components Analysis (PCA) with a varimax rotation to combine the three variables into a single index representing overall quality of life whereby higher values indicate *poorer* health and well-being.

The main independent variables capture various demographic and socioeconomic characteristics that potentially act as mechanisms explaining county-to-county transmission of health and well-being. The selection of mechanisms is based on subdomains established by the CHR model and findings from other studies examining the influence of county-level contextual factors on health and well-being (Arcaya et al., 2016; Diez Rouz and Mair, 2010; Yang et al., 2012, 2015).

The specific subdomains drawn from the CHR include access to clinical health care, quality of clinical health care, poor health behaviors, and social support. Access to health care combines the following variables: percent uninsured and the number of primary care physicians, dentists, mental health providers, and other health care providers per 100,000 population. Clinical health care quality combines the following variables: diabetic monitoring, mammography screening, health care costs, and the number of preventable hospital stays. Poor health behaviors combines measures of tobacco and alcohol use, adult obesity, healthy food environment and physical activity. Social support combines the number of social associations per 10,000 population and individual responses from the BRFSS to a question measuring inadequate social support aggregated up to the county level. The indices were constructed using PCA.

I also included measures of racial segregation and income inequality. I used the non-white/white Dissimilarity index to capture segregation. The Dissimilarity index is a measure of evenness and represents the relative share of group members who would have to exchange places to achieve an even distribution. To capture income inequality, I included the Gini coefficient, which ranges from 0, indicating perfect equality (where everyone receives an equal share of total income), to 1, perfect inequality (where only one recipient or group of recipients receives all the income). I included an index of concentrated disadvantage that combines percent poverty, percent of 25+ year olds with a college degree, median household income, percent of households on public assistance, percent unemployment, and percent of female headed households into a single index using PCA. I also included the number of violent crimes per 100,000 population and log population size.

The CHR provides data for other socioeconomic domains, but these were not included in the model due to high collinearity with the included variables. All social and economic measures were captured temporally prior to the dependent variable. Because of year-to-year variability, I combined years of data. For example, I used the 2015-2017 CHR to measure tobacco use, which uses data from the 2006-12, 2014 and 2015 BRFSS. The years and sources of data for each variable are listed in Table 1.

1.3 Geographic and Migration Weights Matrices

The geographically based spatial weights matrix W_g is based on a queen's case first-order contiguity with a binary weighting scheme. The matrix is nxn, where n represents the number of US counties. Each matrix cell contains a 1 if counties in row j and column k share a border or vertex (point) and 0 otherwise. The matrix is symmetric in that row j and column k has the same value as row k and column j. Other geographic definitions including rook (share a border but not a vertex) and nearest neighbors were tested with results insensitive to the definition.

The second weights matrix is derived from the migration flows between counties. The weights are defined by the number of migrants moving from county to county. In contrast to the geographically based matrix, the migration based matrix W_m is not symmetric. The value in row j and column k represents the number of persons moving from county j to county k. This value is not the same in row k and column k, which represents the number of persons moving from county k and county j. As such, the rows represent the origin or sending counties whereas the columns represent the destination or receiving counties.

I row-standardized both matrices for two reasons. First, this step ensures that the range of cell values is bounded between 0 and 1 for both weight matrices. This is recommended in cases, such as the present case, where nodes are defined by administrative boundaries and thus are susceptible to "edge effects" and other problems that contribute some amount of arbitrary influence to the weight matrix cell values (Anselin 2002; Getis and Aldstadt 2010). Second, row standardizing across both weight matrices makes comparisons across models more valid by constraining the possible values of ρ, the autoregressive term (Bailey and Gatrell 1995). The diagonal contains a 0 in both matrices.

1.4 Modelling Approach

I run a series of spatial dependence models that sequentially reveal the spatial patterning of county-to-county effects. I first run a standard ordinary least squares regression model to examine within county influence of social and economic characteristics. I then run a spatial lag model of the following form

$$Y_i = \rho W Y_i + \beta X_i + \epsilon_i \tag{1}$$

where Y is county overall health well-being, W is the spatial weights matrix defining county connectivity, ρ is the autoregressive coefficient capturing the average effect of a one-unit change in the dependent variable in i's spatially proximate or migration-connected counties (as defined by the given weight matrix) on the dependent variable in the focal county, and X is a set of county-level social and economic characteristics. I run separate models using W_g and W_m.

I then run a spatial durbin model of the following form

$$Y_i = \rho W Y_i + \beta X_i + \gamma W X_i + \epsilon_i \tag{2}$$

where γ represents the average effect of a one-unit change in the independent variable X in county i's spatially proximate or migration-connected counties (as defined by the given weight matrix) on the dependent variable in the focal county. The model is estimated separately for W_g and W_m.

The association between overall health and county-level demographics can be decomposed into direct, indirect, and total effects (LeSage and Pace, 2009). The direct effect refers to the impact of changes in covariates for a given county on changes in the health of that same county (within county impact). The indirect effect refers to the impact of changes in covariates for a neighboring county on the health of the given county (across county impact). The total impact is the sum of the effects of the covariates of the given county and all neighboring counties. All spatial dependence models were estimated using Maximum Likelihood. Whereas a spatial error model is appropriate when one is concerned about unmeasured endogenous effects among connected areas, given that I am interested in exploring two specific contextual processes, the spatial lag and durbin models represent the appropriate choice.

2 Preliminary Results

Table 2 shows results for a standard OLS model and regression models incorporating geographic and migration network lags on the dependent variable. The OLS results reveal that higher levels of concentrated disadvantage, population size, income inequality, poor health behavior, and clinical health care quality are associated with lower health quality. In contrast, greater levels of social support and nonwhite-white segregation are associated with better health quality. Exploratory spatial data analysis reveals significant clustering of the OLS residuals, yielding a Moran's I of 0.519 using a geographically defined spatial weights matrix, which indicates a high correlation between the spatially adjacent counties. The lower Moran's I using the migration network spatial weights matrix (0.080) does not indicate there is no county-to-county influence transmitted through migration flows. Instead, it indicates that the variables included in the model do a good job controlling for these influences.

The spatial lag models control for county-to-county dependencies in geographic and migration connectivity, separately. The Moran's I of the residuals for each model reduces significantly from the OLS model; however, some residual autocorrelation exists in the geographic lag whereas it is near 0 for the migration network lag. The lag coefficient, ρ, is positive and significant in both models, indicating positive influence of overall health of both geographically proximate and migration-connected counties on the overall health of the focal county. The magnitude of the association, however, is much larger in the spatial lag specification, with an autoregressive coefficient nearly twice in size.

The OLS, geographic lag and migration network lag models yield remarkably similar coefficients across the explanatory variables. The exceptions are statistically significant positive effects on clinical health care access in both models but not in the standard OLS specification, indicating an association with poorer health in both the spatial and migration network lag models. Segregation is not significant in the geographic lag model, but remains signifiant and negative in the migration network lag model. Similarly, income inequality is no longer significant in the geographic lag model, but is positive and significant in the migration network lag model.

Table 3 shows results for spatial durbin models. The table is disaggregated by the type of lag, geographic and migration network. Each set of results are separated by direct, indirect, and total impacts. Focusing first on the geographic lag model, the within county (direct) effects largely do not change from the spatial lag results from Table 2. The indirect effects indicate that the poor health behaviors, clinical health care quality, log population size and social support of geographically neighboring counties influence a focal county's health, with increases in the levels of the first three variables leading to poorer health whereas social support is associated with improved health.

Similar to the geographic model, the migration network spatial durbin model yields direct effects that are similar to the migration network lag model. The indirect effects indicate that the log population size and poor health behaviors of migration-connected counties influence county health and well-being. In both cases, higher levels are associated with poorer health. These results are similar to those found in the geographic lag model, but the effect sizes are much larger. I also find that social support influences county health through migration flows with the coefficient size similar across the geographic and migration network models. There are three major differences between the two models. Unlike for geographically proximate neighbors, the average health care access in migration network connected counties have no association. I also find that concentrated disadvantage influences county health for migration network connected counties but not for geographically adjacent counties. In this case, higher levels of concentrated disadvantage in migration-connected areas are beneficial to a county's health.

In summary, I found that the average overall health in geographically adjacent and migration connected counties influence the overall health in a focal county. That is, a county's health is influenced not just by the overall health in neighboring counties, as has been extensively established in prior work, but also the overall health in counties connected via migration flows. I also found both similar and differential effects of various social and economic characteristics via geographic and migration networks. Both specifications indicate the influence of population size, poor health behaviors, and social support. However, clinical health care quality in nearby counties but not in migration connected counties impact overall health. In contrast, concentrated disadvantage influences county health via migration flows but not geographic connectedness. These results indicate similar and differential ways through which social and economic characteristics diffuse to impact county health.

3 Next Steps

I will extend the current version of this paper by (1) providing descriptive results showing how counties are connected via migration streams by examining the distribution of migration ties across counties. The objective of this stage of analysis is to uncover the structure of the migration flow network and how it differs from a network based on spatial adjacency; and (2) examining whether I can incorporate geographic and migration network lags into single lag and durbin models.

4 References

Anselin, L. (2002). Under the hood: Issues in the specification and interpretation of spatial regression models. Agricultural economics, 27(3), 247-267.

Arcaya, M. C., Tucker-Seeley, R. D., Kim, R., Schnake-Mahl, A., So, M., & Subramanian, S. V. (2016). Research on neighborhood effects on health in the United States: A systematic review of study characteristics. Social Science & Medicine, 168, 16-29.

Bailey, T. C., & Gatrell, A. C. (1995). Interactive spatial data analysis (Vol. 413). Essex: Longman Scientific & Technical.

Brazil, N. (2017). Spatial variation in the Hispanic Paradox: Mortality rates in new and established hispanic US destinations. Population, space and place, 23(1), e1968.

Brown, D. L. (2002). Migration and community: Social networks in a multilevel world. Rural Sociology, 67(1), 1-23.

Browning, C. R., Dietz, R. D., & Feinberg, S. L. (2004). The paradox of social organization: Networks, collective efficacy, and violent crime in urban neighborhoods. Social Forces, 83(2), 503-534.

Christakis, N. A., & Fowler, J. H. (2008). The collective dynamics of smoking in a large social network. New England journal of medicine, 358(21), 2249-2258.

Haynie, D. L. (2001). Delinquent peers revisited: Does network structure matter?. American journal of sociology, 106(4), 1013-1057.

Hipp, J. R., & Perrin, A. J. (2009). The simultaneous effect of social distance and physical distance on the formation of neighborhood ties. City & Community, 8(1), 5-25.

Getis, A., & Aldstadt, J. (2004). Constructing the spatial weights matrix using a local statistic. Geographical analysis, 36(2), 90-104.

Leenders, R. T. A. (2002). Modeling social influence through network autocorrelation: constructing the weight matrix. Social networks, 24(1), 21-47.

LeSage, J., & Pace, R. K. (2009). Introduction to spatial econometrics. Chapman and Hall/CRC.

Mears, D. P., & Bhati, A. S. (2006). No community is an island: The effects of resource deprivation on urban violence in spatially and socially proximate communities. Criminology, 44(3), 509-548.

Oakes, J. M., Andrade, K. E., Biyoow, I. M., & Cowan, L. T. (2015). Twenty years of neighborhood effect research: an assessment. Current epidemiology reports, 2(1), 80-87.

Radil, S. M., Flint, C., & Tita, G. E. (2010). Spatializing social networks: Using social network analysis to investigate geographies of gang rivalry, territoriality, and violence in Los Angeles. Annals of the Association of American Geographers, 100(2), 307-326.

Remington, P. L., & Booske, B. C. (2011). Measuring the health of communitieshow and why?. Journal of Public Health Management and Practice, 17(5), 397-400.

Roux, A. V. D., & Mair, C. (2010). Neighborhoods and health. Annals of the New York Academy of Sciences, 1186(1), 125-145.

Smith, K. P., & Christakis, N. A. (2008). Social networks and health. Annu. Rev.

Sociol, 34, 405-429.

Sparks, P. J., & Sparks, C. S. (2010). An application of spatially autoregressive models to the study of US county mortality rates. Population, Space and Place, 16(6), 465-481.

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.

Tobler, W. R. 1970. A computer movie simulating urban growth in the Detroit region. Economic Geography 46:23440

Valente, T. W. (1996). Social network thresholds in the diffusion of innovations. Social networks, 18(1), 69-89.

Wellman, Barry. (Ed.). 1999. Networks in the Global Village. Boulder, CO: Westview.

Yang, T. C., Chen, V. Y. J., Shoff, C., & Matthews, S. A. (2012). Using quantile regression to examine the effects of inequality across the mortality distribution in the US counties. Social science & medicine, 74(12), 1900-1910.

Yang, T. C., Noah, A. J., & Shoff, C. (2015). Exploring geographic variation in US mortality rates using a spatial Durbin approach. Population, space and place, 21(1), 18-37.

5 Tables

Table 1: Variable data sources and time period

Domain	Variable	Year	Data Source
Quality of life	Poor or fair health	2016	BRFSS
	Poor physical health days	2016	BRFSS
	Poor mental health days	2016	BRFSS
	Percent 65 and below uninsured	2011-2015	SAHIE
Access to healthcare	Primary care physicians per 100,000	2011-2015	AHRF/AMA
	Dentists per 100,00	2012-2015	AHRF/NPI
	Mental health providers per 100,000	2013-2015	CMS/NPI
	Other primary care providers per 100,00	2013-2015	CMS/NPI
Healthcare quality	Percent diabetic Medicare 65-75 years old that receive monitoring	2011-2014	Dartmouth
	Percent female Medicare 67-69 that receive mammography screening	2011-2014	Dartmouth
	Amount of price-adjusted Medicare reimbursements per enrollee	2011-2015	Dartmouth
	Number of hospital stays per 1,000 Medicare enrollees	2011-2015	Dartmouth
Health behaviors	Percent smokes every day or most days	2006-12, 2014, 2015	BRFSS
	Percent binge drinking	2014-2015	BRFSS
	Percent who live close to a location for physical activity	2010, 2012-2014	Multiple
	Obesity rate	2011-2014	CDC
	Percent 20+ reporting no leisure-time physical activity	2011-2014	CDC
Social support	Number of social associations per 10,000	2012-2015	CBP
	Inadequate social support	2005-2010	BRFSS
Concentrated disadvantage	Percent poverty	2011-2015	ACS
	Percent 25+ with college degree	2011-2015	ACS
	Median household income	2011-2015	ACS
	Percent of households on public assistance	2011-2015	ACS
	Percent unemployed	2011-2015	ACS
	Percent of female-headed households	2011-2015	ACS
	Nonwhite/white Dissimilarity Index	2011-2015	ACS
	Gini index	2011-2015	ACS
	Log population size	2011-2015	ACS
	Violent crimes per 100,000		UCR

BRFSS: Behavioral Risk Factor Surveillance System; SAHIE: Small Area Health Insurance Estimates; AHRF: Area Health Resource File; AMA: American Medical Association; NPI: National Provider Identification file; CMS: Centers for Medicare and Medicaid Services; Dartmouth: Atlas of Health Care; Multiple: Business Analyst, Delorme map data, ESRI, & US Census Tigerline Files; CDC: Centers for Disease Control and Prevention Diabetes Interactive Atlas; CBP: County Business Patterns; ACS: American Community Survey; UCR: Uniform crime reporting FBI

Table 2: Ordinary Least Squares (OLS) and Spatial Lag Models of County Health in 2016

	OLS	Geographic Lag	Migration Network Lag
Log population size	0.073*** (0.008)	0.030*** (0.006)	0.058*** (0.008)
Gini index	1.508*** (0.295)	0.270 (0.225)	1.248*** (0.288)
White-Non-white Segregation	-0.004*** (0.001)	-0.001 (0.001)	-0.003*** (0.001)
Violent crime rate	0.0001 (0.0001)	0.0001 (0.0000)	0.0001 (0.0001)
Poor health behavior	0.352*** (0.013)	0.158*** (0.011)	0.321*** (0.013)
Clinical health care quality	0.134*** (0.011)	0.066*** (0.008)	0.131*** (0.011)
Clinical health care access	0.020 (0.010)	0.027*** (0.008)	0.024* (0.010)
Concentrated disadvantage	0.440*** (0.014)	0.375*** (0.011)	0.447*** (0.013)
Social support	-0.124*** (0.011)	-0.051*** (0.008)	-0.118*** (0.011)
Intercept	-1.326*** (0.148)	-0.416*** (0.113)	-1.080*** (0.145)
$\rho\ W_g$		0.514*** (0.011)	
$\rho\ W_m$			0.261*** (0.022)
AIC	4297.5	2687.9	4152.3
Moran's I W_g	0.519	0.159	
Moran's I W_m	0.08		0.026

***p $\leq$ 0.001 **p $\leq$ 0.01 *p $\leq$ 0.05

Table 3: Spatial Durbin Models of County Health in 2016

	Geographic Lag			Migration Network Lag		
	Direct	Indirect	Total	Direct	Indirect	Total
Log population size	0.002	0.100*	0.102	0.059***	0.417*	0.476*
Gini index	1.096***	-1.977	-0.881	1.245***	-1.343	-0.098
White-Non-white Segregation	-0.0004	-0.010	-0.010	-0.003***	-0.028	-0.032
Violent crime rate	0.0000	0.0000	0.0000	0.0001	0.001	0.001
Poor health behavior	0.214***	0.258***	0.472***	0.319***	0.558*	0.877***
Clinical health care quality	0.056***	0.112*	0.167**	0.129***	0.255	0.385*
Clinical health care access	0.022***	0.157	0.179*	0.021*	0.152	0.173
Concentrated disadvantage	0.448**	0.070	0.518***	0.449***	-0.223*	0.226
Social support	-0.043***	-0.167*	-0.211*	-0.114***	-0.197*	-0.311
ρ W_g		0.899***				
ρ W_m					0.831***	
AIC		2008.1			4095.7	
Moran's I W_g		-0.008				
Moran's I W_m					-0.001	

***$p \leq 0.001$ **$p \leq 0.01$ *$p \leq 0.05$