Social Science & Medicine xxx (xxxx) xxx

Contents lists available at ScienceDirect

Social Science & Medicine

journal homepage: http://www.elsevier.com/locate/socscimed

The multidimensional clustering of health and its ecological risk factors

Noli Brazil

University of California, Davis, Department of Human Ecology, One Shields Ave, Davis, CA, 95616, USA

ARTICLE INFO

Keywords:
Spatial analysis
Social determinants of health
Social networks
Migration
Syndemic

ABSTRACT

A diverse set of research has examined the ways in which population-level health and its ecological risk factors are embedded within self-reinforcing structures. Syndemic theory, for example, focuses on the co-occurrence of multiple diseases, whereas the spatial diffusion literature highlights the concentration of poor health among communities sharing geographic boundaries. This study combines these related but disciplinarily-isolated perspectives to examine the clustering of population-level health and its determinants across four dimensions: co-occurrence, spatial, temporal, and social network. Using data on U.S. county-level health outcomes and health factors from the Robert Wood Johnson Foundation's County Health Rankings, this study estimates associations between health outcomes within communities and the co-occurrence of community-level factors theorized to influence ecological health. Not only do health outcomes and their ecological risk factors cluster within counties, but also between geographically adjacent counties and counties connected via migration network pathways. Moreover, the self-reinforcing structures uncovered across the co-occurrence, spatial and network dimensions persist over time, and this clustering has consequences on county health and well-being. Rather than adopting the perspective that either health and its community-level factors should be broadly targeted and detached from local context or communities are different, have unique needs and thus should be treated in isolation, the approach advanced in this study identifies shared vulnerabilities in a way that allows for the development of knowledge networks between communities dealing with similar issues.

1. Introduction

In 2015, for the first time in more than two decades, life expectancy in the United States declined (Xu et al., 2016). This decline continued until 2018 when life expectancy only slightly increased. The decline in life expectancy was unequally distributed across the country as counties with the highest life expectancies have continued to increase life expectancy, while those with the lowest have plateaued (Dwyer-Lindgren et al., 2017; Vierboom et al., 2019). There is growing recognition in research, policy and practice that adequately addressing this rising health inequality will require approaches that move beyond standard interventions that focus on a single health outcome and household or individual-level determinants such as medical care access and quality (Zajacova and Montez, 2017). An increasingly popular approach is to target multiple co-occurring health conditions in a single intervention. This approach draws from syndemic theory, which highlights how synergistic interactions among psychosocial and biological conditions underlie patterns of disease clustering (Singer, 1996; Tsai, 2018; Tsai and Venkataramani, 2016). It is not a single disease or multiple diseases acting independently of one another that is solely contributing to stagnating or declining life expectancies in certain communities, but the co-occurrence or synergistic interaction of several diseases (Meader et al., 2016; Tsai, 2018).

Syndemic theory points to large-scale social, political, economic, and ecological factors giving rise to disease co-occurrence and clustering (Singer et al., 2017). The perspective that contextual characteristics influence health outcomes is not new. For example, the social determinants of health (SDH) framework argues that health inequalities cannot be attributed to differences in individual characteristics alone; the upstream contextual determinants that reflect the economic and social resources and opportunities influencing residents' access to health-promoting living and healthy choices strongly shape ecological health and well-being (Marmot, 2005; Bambra et al., 2010). Other ecological factors, such as access to clinical care, community-level health behaviors, and the physical environment, also play a role in shaping health above and beyond individual characteristics (Hood et al., 2016). The SDH framework and similar perspectives coupled with syndemic theory predict that ecological factors have both independent and interactive effects on health. That is, both health and its ecological risk factors co-occur, and this co-occurrence amplifies the disease burden

E-mail address: nbrazil@ucdavis.edu.

https://doi.org/10.1016/j.socscimed.2021.113772
Received in revised form 7 February 2021; Accepted 9 February 2021
Available online 16 February 2021
0277-9536/© 2021 Elsevier Ltd. All rights reserved.

experienced by disadvantaged communities.

The co-occurrence of risk factors and health outcomes is only one type of clustering. Drawing from a wide variety of social science perspectives, including syndemic theory, the SDH framework, and social network and spatial perspectives, I argue that community health and its ecological risk factors are embedded within several self-reinforcing structures. Without accounting for clustering across these multiple structures, we are underestimating the health burden experienced in certain areas of the country. Using county-level demographic data from the American Community Survey (ACS) and health outcomes and their ecological risk factors from the 2015 and 2018 Robert Wood Johnson Foundation's County Health Rankings (CHR), I examine the magnitude and extent of clustering in health outcomes and ecological risk factors across four dimensions: co-occurrence, spatial, temporal, and social network. The study's objective is to demonstrate that community-level health and its ecological risk factors are clustered across these dimensions, and thus not accounting for these various forms of embeddedness underestimates a community's disease burden.

1.1. Conceptual model

Co-occurrence and spatial, temporal, and social network clustering are represented, without reference to specific contextual factors, geographic context and health outcomes, in Fig. 1. The gray polygons represent geographic areas such as counties, with arrows depicting associations within and between geographic areas. Arrows colored in black represent the pathways captured by each dimension of clustering. Fig. 1a depicts variable clustering or co-occurrence, which is the concomitant association between multiple health outcomes and its risk factors in a given place and time. Co-occurrence research has found evidence of a synergistic effect of health conditions; this research, however, has largely focused on studying individuals rather than populations or communities and on the co-occurrence of health conditions but not their risk factors, and has neglected the spatial and network dimensions of co-occurrence. (Tsai, 2018; Tsai et al., 2017).

Fig. 1b depicts temporal clustering, which measures the persistence of health outcomes and health risk factors across time. Significant attention has been paid towards the transmission of individual health and its risk factors within and between generations (Wickrama, 1999) and the enduring effects of contextual risk factors, such as neighborhood poverty and violence, on individual health across the life course (Brazil and Clark, 2017; Sharkey, 2008). The current study examines how advantages and disadvantages in ecological health and its determinants persist over time. This dimension of clustering directs attention to the trajectory of population health. Appropriate interventions in counties exhibiting poor health over a longer period of time will differ from interventions in counties experiencing either poor but improving health or good but declining health.

Fig. 1c depicts spatial clustering, which measures the concentration of health outcomes and their ecological risk factors within a geographically bounded group of communities. The spatial clustering literature has demonstrated that health outcomes in areas as small as census tracts and as large as counties are associated with health outcomes in geographically adjacent areas. Studies have demonstrated that a variety of health outcomes follow this geographic spillover process, including mortality rates (Brazil 2017; Sparks and Sparks, 2010; Yang et al., 2015), diabetes (Myers et al., 2017), mental health (Gruebner et al., 2015; Yang 2019), heart disease (Kramer et al., 2017), and composite indices of overall health and well-being (Tabb et al., 2018). Not only can health outcomes diffuse across geographic boundaries, but also their ecological social and demographic determinants. For example, Yang et al. (2015) found that social capital in a specific county is negatively associated with the mortality rates in that county and its neighboring counties. High spatial clustering suggests that health outcomes and its determinants diffuse or spill over to geographically proximate places (Tabb et al., 2018; Yang et al., 2015). Consequently, an area of poor

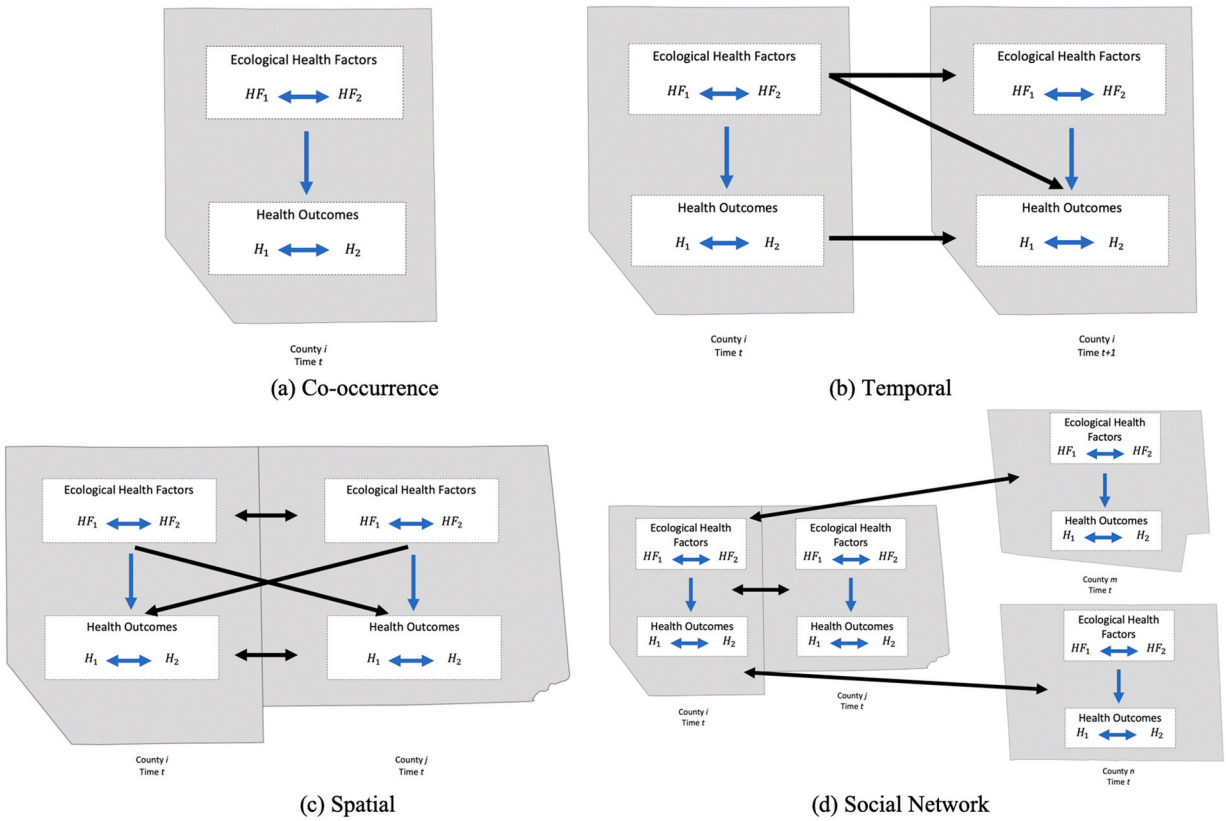


Fig. 1. Conceptual model of the clustering of health (H) and its ecological risk factors (HF) across co-occurring, temporal, spatial and network dimensions.

health becomes embedded within a larger geographic cluster of poor health communities with self-reinforcing spatial structures. In order to dismantle these structures, interventions should not only target a community's health outcomes and their determinants, but also the outcomes and risk factors of its geographic neighbors

The spatial adjacency literature treats diffusion as a geographically bounded process, similar to an infection spreading from a localized point source. Here, geographic proximity is the primary determinant of the ties that foster the diffusion of health and its determinants. Strong empirical evidence indicates that the diffusion of health is spatially structured, even at varying levels of geographic aggregation; thus, this approach has long been considered an adequate explanatory model (Matthews and Yang, 2013). But, a separate perspective, social network theory, predicts that diffusion is not strictly a spatially bounded process (Papachristos and Bastomski, 2018). Fig. 1d depicts network clustering, which measures the synergistic interaction of health and community-level factors between communities connected via economic, political and demographic pathways that are not completely governed by geographic proximity. An established literature has demonstrated the spread of health conditions and their risk factors across a social network defined not simply by direct kin ties, but also by ties extending to several degrees of separation (Christakis and Fowler, 2007; Umberson and Montez, 2010; Papachristos and Bastomski, 2018). What counts is a person's placement in social space, not geographical space. Only recently have social scientists extended social network theory to connect ecological units, with the majority of these applications found in criminology (Tita and Radil, 2010; Bastomski et al., 2017; Graif et al., 2019). The present study examines social network clustering via a particular pathway connecting communities both near and far: migration flows.

The justification for examining migration flows stems in part from the observation that health outcomes and health factors are not airborne phenomena that disperse across boundaries. Rather, people move, taking their health and behaviors with them. From a networked perspective, the choice to move to a certain county is shaped not just by individual happenstance or spatial proximity, but also by meso-level institutional ties or meaning frameworks (Fawcett 1989), as well as social distance between areas. Individual residential mobility choices are also influenced by larger structures that are shaped by meso-level processes such as segregation, gentrification, or housing policies (Clark and Brazil, 2019; Clark and Maas, 2015). Such a proposition follows work on how migration creates or disrupts networks between local communities within a city (Sampson 2012). It is also motivated by the growing literature on migration network theories that explain flows between communities from different countries via the establishment of social and institutional ties (Danchev and Porter 2018; DeWaard and Ha, 2019; Leal et al., 2019). According to these theories, migration from an origin community to a foreign destination community is established by broader forces making those flows possible in the first place. As these flows continue, origin and destination become socially and economically linked, especially if counterflows exist and migrants maintain ties to origin communities. The existence of global interconnectedness and mutual migrant streams is likely to generate "multilateral migration structures that are irreducible to preexistent geographic boundaries or independent migration exchanges" between pairs of communities (Danchev and Porter 2018: 7). These sorts of interactions give the network its form and feed back into the ways it affects both communities and individuals.

The overarching theoretical position of this study is that a complete understanding of a population's disease burden needs to incorporate the clustering of health and its ecological risk factors across multiple dimensions. Although several research domains have articulated some of these dimensions, they often do not speak to one another despite strong similarities in their underlying empirical and theoretical underpinnings. Consequently, we are left with an incomplete picture of health disadvantage. This study brings together these interrelated research domains by testing for evidence of clustering across the multiple pathways through which ecological health and its ecological factors are connected. The study's primary aim is to demonstrate that community-level health is dependent on multiple forms of embeddedness either not previously accounted for or separately examined by the current health literature.

2. Data and methods

2.1. Data

In selecting health outcomes and their ecological risk factors, I adopted CHR's framework, a widely accepted model given its comprehensiveness, reliance on publicly available data, and its theoretical grounding via a rigorous review of the population health literature (Remington et al., 2015). The model describes a holistic view of population health, highlighting multiple ecological factors including SDH and their relative contributions to length of life and quality of life. The CHR model identifies 13 health risk factors organized under four broad categories: Health behaviors, which captures tobacco use, diet and exercise, alcohol and drug use, and sexual activity; Clinical care, which captures access to and quality of health care; Social and economic factors, which includes education, employment, income, family and social support, and community safety; and physical environment, which includes air and water quality and housing and transit. Each sub-domain is represented by a range of measures, which I standardized and averaged to obtain 13 sub-domain indices.

The CHR uses two health outcomes, length of life, measured by the potential life lost before age 75 per 100,000 population, and quality of life, measured by the percentage of adults reporting fair or poor health, average number of physically unhealthy days reported in the past 30 days, average number of mentally unhealthy days reported in the past 30 days, and percentage of live births with low birthweight. I standardized these variables and took the average to obtain an overall health index. Because syndemic theory centers on the notion that multiple adverse conditions interact synergistically, I also examined other health outcomes, specifically diabetes prevalence (percentage of adults aged 20 and above with diagnosed diabetes), one of the leading noncommunicable diseases in years of life lost and disability adjusted life years in the United States and a disease typically examined by syndemic research because of its high and increasing prevalence (Menke et al., 2015) and strong co-occurrence with other diseases (Mendenhall, 2016), and premature age-adjusted mortality (number of deaths among residents under age 75 per 100,000 population), which is used as a cumulative measure of health.

I analyzed 2018 CHR health data for counties in the contiguous United States (N = 3108), with measures reflecting the most current year that data are available. I used data from the 2015 CHR to capture past ecological characteristics. The 2015 CHR data are generally measured 2–5 years prior to the 2018 CHR values, with some minimal overlap. Several variables are estimated from state-level data from the Behavioral Risk Factor Surveillance System using Bayesian multilevel modeling techniques Barker et al. (2013). The years and sources of data for each variable are listed in the Appendix Table S1. Descriptive statistics for each variable are provided in Appendix Table S2.

2.2. Statistical analysis

In this study, I used descriptive correlations and spatial and network methods to measure clustering across the four dimensions and model the association between clustering and community health and well-being. I did not attempt to measure causal pathways; instead, I focused on descriptively establishing baseline levels of clustering and establishing their association with health outcomes. I measured co-occurrence two ways. First, I calculated Pearson correlation coefficients of health outcomes and their determinants. Next, I performed an area classification, which is a cluster analysis procedure that classifies areas into groups on

the basis of the similarity of characteristics of selected features within them. Specifically, I conducted a geodemographic analysis, which is used as a means of multivariate data reduction to classify people according to where they live (Singleton and Spielman, 2014). Geodemographic analysis is a clustering procedure that organizes each area (often specified at a small geographic scale) into groups based upon the overall similarities concealing within those multivariate attributes which they share. In addition to CHR's 13 sub-domains, I included in the geodemographic analysis demographic variables not included in CHR's framework that both capture important geographic variation in the demographic profiles of counties and may help explain county-level variation in health conditions (Delmelle, 2019; Wallace et al., 2019). These variables are educational attainment (percent without a high school degree and percent with a college degree), race/ethnicity (percent American Indian, percent Asian, percent white, percent black, and percent Hispanic), age composition (percent below 18 years old, percent between 25 and 34 years old, and percent 65 years old and over) and percent rural. The demographic variables were downloaded from the 2011–2015 ACS. The age groups chosen represent distinct periods in the life course (childhood and adolescence, young adulthood and senior and older ages) that are associated with different health conditions, risk factors and lifestyles that help differentiate community types such as retirement communities catered to seniors and urban amenity-rich communities catered to young professionals (Delmelle, 2019; Grubesic et al., 2014). Racial/ethnic composition was included because of the strong degree of racial and ethnic residential concentration in specific parts of the country, such as Native American communities in the West and Southwest (Norris et al., 2012) and traditional Hispanic and Asian destinations in the West, Southwest, Sun Belt and the East (Lichter and Johnson, 2009). Moreover, differentiating between community types by their health risk factors must consider the interaction between race/-ethnicity and levels of risk. For example, communities of black and Hispanic populations tend to be more disadvantaged than those of comparable white populations, and thus typically have lower access to clinical care and poorer housing, transit, and environmental quality (Williams and Sternthal, 2010). However, poor health risk factors also exist in disadvantaged white communities, such as those in the rural South (Fenelon, 2013), and historically disinvested Native American communities (Barnes et al., 2010). Similarly, minority communities are not homogenously high-risk, but some have positive health factors, such as those documented in ethnic enclaves (Markides and Coreil, 1986; Tam, 2019).

I used hierarchical clustering with Euclidean distance to measure similarity and Ward's minimum variance method to minimize within-group differences and maximize between-group differences. Rather than relying on one metric to determine the optimal number of k clusters, I used the complete set of evaluation metrics provided by the R package NbClust (Charrad et al., 2014). The package provides 30 indices, including metrics such as within and between sum of squares and Gap statistics. These indices combine information about intracluster compactness and intercluster isolation, as well as other factors, such as geometric or statistical properties of the data, and dissimilarity and similarity measures, with the aim of finding the k that best separates observations into distinct classes. The procedure provides the values of k by the number of indices indicating their optimality. The optimal k ranges from 6 to 10, with no one value representing the clear majority, and thus I used the median number of this range $k = 8$.

To measure temporal clustering, I examined Pearson correlation coefficients of health outcomes and their ecological risk factors across two time points. To measure spatial clustering, I examined the Global Moran's I, the spatial equivalent to the Pearson correlation, for each health outcome and health factors. The Moran's I is based on an n x n weights matrix W_g that defines geographic neighbors, in this case Queen contiguity, whereby each matrix cell contains a 1 if counties in row j and column k share a border or vertex (point) and 0 otherwise. The Moran's I is a standard measure of spatial autocorrelation with values ranging from −1 to 1. A Moran's I equal to 0 indicates a random spatial pattern with no spatial autocorrelation. Positive values suggest spatially clustered patterns in adjacent areas, whereas negative values indicate that samples reveal very different values from the neighboring ones.

I also used Moran's I to measure migration network clustering. Moran's I is a valid test of network dependence (Lee and Ogburn, 2020) and has been used as a measure of clustering in applications combining network and spatial perspectives (Bastomski et al., 2017; Papachristos and Bastomski, 2018). Other measures of network dependence, segregation or homophily have been proposed, but are for categorical variables based on group or class affiliation such as race/ethnicity (Bojanowski and Corten, 2014). Moreover, measuring Moran's I for both spatial and network allows for a standardized comparison of clustering across models. The weights in the migration network matrix W_m are defined by the number of migrants moving from county to county. The rows represent the destination or receiving counties and the columns represent the origin or sending counties. I used U.S. Census County-to-County Migration Flows data to measure migration flows for every county pair in the contiguous 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 and where they lived 1 year prior. I row-standardized both spatial and network based matrices such that the sum of values across a row equals to 1. The value in cell (j, k) in the row-standardized weights matrix W_m represents the proportion of individuals migrating to county j (row) that originated from county k (column).

Because propinquity confounds network connectivity, I also constructed a weights matrix W_{gm} that accounts for both migrant and geographic connectivity. Rather than the absolute number of individuals migrating from county i to county j as captured in the weights matrix W_m, I used the number of migrants weighted by the inverse distance between the two counties, where distance is measured by Euclidean distance between county centroids. This modifies W_m by decreasing the influence of counties that are farther away. Similar to W_m and W_g, I row standardized W_{gm}.

Having established the levels of clustering at each dimension, regression models estimating the association between a county's health and the health and ecological risk factors of its temporal, spatial and migration network neighbors were run. I included all 13 CHR sub-domains in the models in order to preserve the framework that guided the selection of variables in the clustering analysis. Multicollinearity was not present as variance inflation factors from ordinary least squares (OLS) regressions were below standard acceptable levels (i.e. below 5). To estimate the influence of temporal clustering, I estimated an OLS regression of the following form

$$Y_{it} = \alpha + \beta X_{it} + \rho Y_{it-n} + \gamma X_{it-n} + \theta C_{it} + \varepsilon_{it}$$

where Y_{it} is health outcome for county i at time t, X_{it} is the set of ecological health risk factors measured at time t, Y_{it-n} is health outcome measured at time t-n, X_{it-n} is the set of ecological health factors measured at time t-n, C_{it} is a set of control variables (% Hispanic, % rural, log population size, and in- and out-migration rates), and ε_{it} is a random error term. Health and ecological factors measured at time t-n is defined as the temporal lags (the association between current health and health and its risk factors from a prior period).

To examine the association between health and spatial and network clustering, Spatial Durbin regression models of the following form were estimated using 2018 CHR data (Anselin, 1988)

$$Y_i = \alpha + \beta X_i + \rho WY + \gamma WX_i + \theta C_i + \varepsilon_i$$

where Y_i is the health outcome for county i, X_i is the set of ecological health factors, W is the weights matrix, W_g, W_m or W_{gm}, C_i is a set of control variables, and ε_i is a random error term. The coefficient ρ is the health lag effect, which estimates the association between a county's

health and the average health in that county's spatial- or migration-connected counties. The coefficients β and γ represent the direct and indirect effects of X_i (LeSage and Pace, 2009). The direct effect refers to the association between changes in covariates for a given county and changes in the health of that same county. The indirect or ecological risk lag effect refers to the association between a county's health and the average ecological risk characteristics in that county's geographically adjacent (spatial) or migration sending (network) network. All analyses were conducted in R using the packages spdep and spatialreg. All models were estimated using maximum likelihood. All maps were constructed using the Geographic Information Systems software QGIS version 3.16.0.

3. Results

Fig. 2 visually summarizes the co-occurrence of CHR's 3 health outcomes and 13 health risk factors through a correlogram. I transformed all ecological factors such that greater values indicate poorer conditions. Darker colors indicate higher correlations and only coefficient values with p-values less than 0.01 are shown. With the exception of air and water quality and access to clinical care, the plot broadly shows strong statistically significant correlations amongst the ecological risk factors. There is also strong co-occurrence between the three health outcomes: correlation values of −0.87, 0.68 and −0.71 for mortality and overall health, mortality and diabetes prevalence, and diabetes prevalence and overall health, respectively.

An alternative approach to measuring co-occurrence is to examine how groups of variables hang together. Fig. 3a displays the locations of the 8 clusters produced through a geodemographic classification using CHR's 13 health risk factors and basic demographic characteristics. The clusters are labelled according to their most distinguishing features (see Appendix Table S2 for cluster mean values of the characteristics used in the classification and Appendix Table S3 for cluster descriptions). Fig. 3b through 3d map the three health outcomes. A few important findings

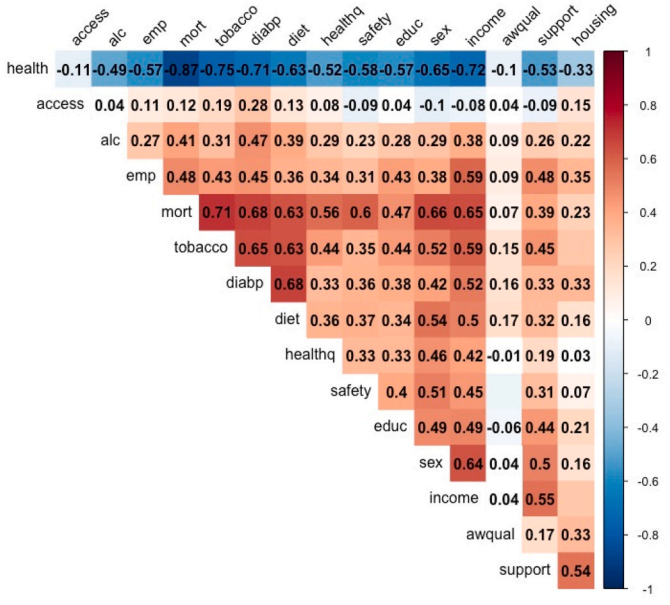


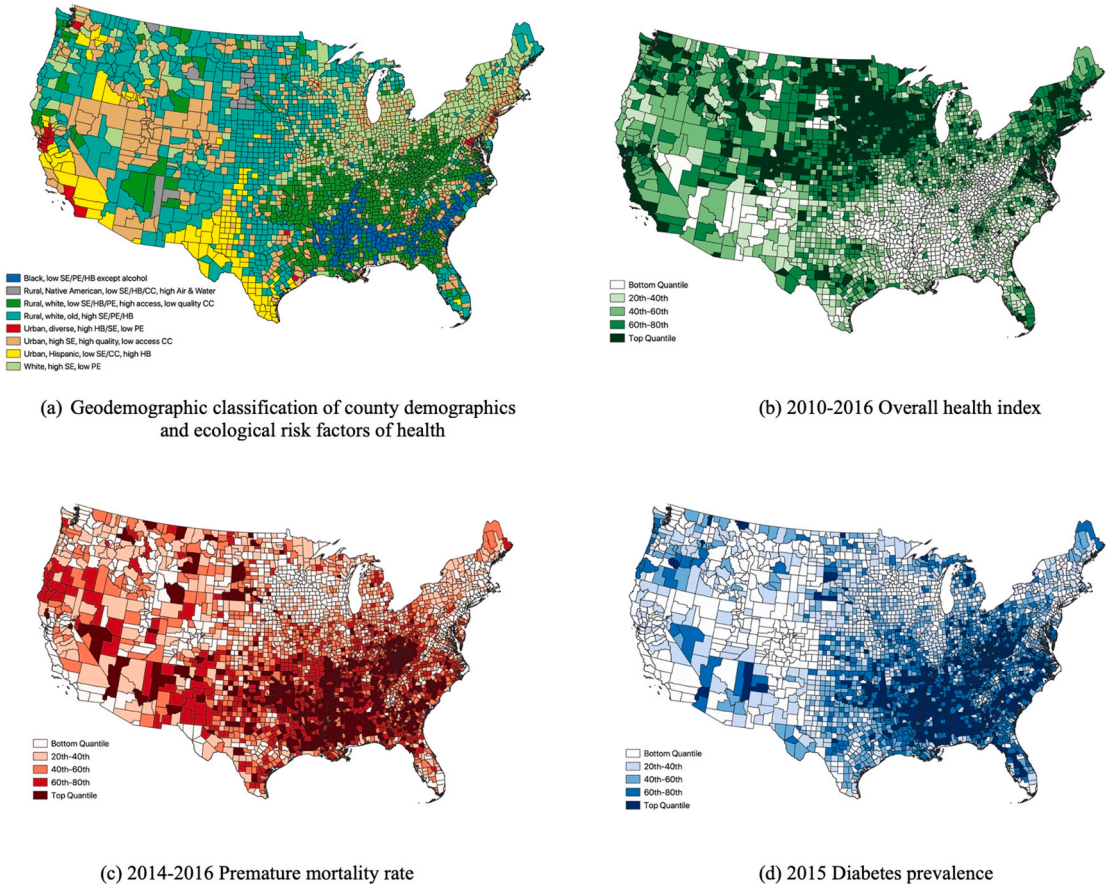
Fig. 2. Correlogram of health outcomes and ecological factors. Pearson correlation values statistically significant at the $p < 0.01$ level are shown. Abbreviations: health: overall health; access: access to clinical care; alc: alcohol and drug use; emp: unemployment rate; mort: premature mortality rate; tobacco: tobacco use; diabp: diabetes prevalence; diet: diet and exercise; healthq: quality of clinical care; safety: community safety; educ: education; sex: sexual activity; income: poverty and income; awqual: air and water quality; support: family and social support; housing: housing and transit. Appendix table S1 provides descriptions of each variable.

emerge from the classification. First, separation is largely based on the interactions between a county's health risk factors and its racial and ethnic composition and rurality. That is, it is not just the variation in how health risk factors co-occur that differentiates counties, but how this co-occurrence interacts with a county's racial and ethnic composition. Furthermore, counties are not universally categorized into health enhancing or diminishing classes, but are separated by groups of characteristics that often covary in opposite directions. For example, the analysis identified a group of racially/ethnically diverse urban counties with positive health behaviors but poor physical environments. County characteristics may even point in opposite directions within each broad domain, such as low clinical care access, but high clinical care quality in urban, diverse counties. Second, the classes spatially cluster, and geographically align with the spatial patterning of the health outcomes. For example, Fig. 3a shows the strong spatial clustering of high black population, negative health factor counties in the South. This cluster visually aligns with the cluster of high mortality, high diabetes, and poor overall health counties shown in Fig. 3b–d. Running a geodemographic classification of the three health outcomes yielded an optimal cluster size $k = 3$, with the three classifications reflecting high, mid and low values across the three outcomes. Because the classification identified categories indicating strong correlation across the outcomes, I do not report the results.

Having established strong co-occurrence amongst health outcomes and their ecological factors, I present temporal, geographic and migration network correlations in Table 1. The first column of values presents Pearson correlation coefficients of health outcomes and health factors over an approximately 5-year period (see Appendix Table S1 for specific years). I find nearly all correlations are greater than 0.50. Moreover, not only do individual determinants correlate over time, but also their co-occurrence. Table 2 shows the percent of counties by geodemographic cluster type in two separate time periods (See Appendix Table S3 for crosswalk between cluster number and distinguishing characteristics). The diagonal of this transition matrix, which represents the percent of counties classified in the same cluster type in both years, shows that the significant majority of counties are grouped together based on their health risk factors across time.

The rest of Table 1 shows correlations of health and their ecological risk factors organized by CHR's four umbrella categories across spatial, where spatial is represented as counties sharing the same borders or vertices, migration network, where network ties are based on counties connected via migration flows, and migration ties inversely weighted by distance. For all network structures, I calculated the global Moran's I for two time periods. Counties exhibit greater correlation across geographic than network pathways for all determinants and health outcomes. However, the correlations are generally high for both, ranging from 0.18 to 0.66 and 0.16 to 0.49 for geographic and migration network clustering, respectively, and are similar across time. The correlations for the combined network generally lie somewhere in between the strictly spatial and migration based correlations.

Having established that health outcomes and their ecological risk factors cluster across temporal, geographic and network pathways, I model the consequences of this clustering on county health. Specifically, I examine the average association between a county's health and the health outcomes and their risk factors in that county's temporal, spatial and network neighbors. Results from a series of temporal, spatial and network lag regression models are presented in Tables 3–5. These models estimate three effects on health: (1) the temporal, spatial, network and combined spatial and network lag of health; (2) the direct effects of a county's ecological factors; and (3) the indirect or lag effects of health risk factors for a county's temporal, spatial and network neighbors. The temporal lag of health is a county's health measured at a prior time period. The spatial and network lag of health represent the average health levels in a county's geographically adjacent neighbors and migrant sending counties, respectively. The direct effect measures the association between a county's current ecological risk factors and its



(a) Geodemographic classification of county demographics and ecological risk factors of health

(b) 2010-2016 Overall health index

(c) 2014-2016 Premature mortality rate

(d) 2015 Diabetes prevalence

Fig. 3. Maps of U.S. county health outcomes and the geodemographic classification of health risk factors. The geodemographic classification of county demographics and ecological risk factors of health is based on the 2018 Community Health Rankings and 2011–2015 American Community Survey (SE: Social and Economic, which combines Education, Employment, Income, Family & Social Support, and Community safety domains, PE: Physical Environment, which combines Air & Water Quality and Housing & Transit domains, HB: Health Behaviors, which combines Tobacco use, Diet & Exercise, Alcohol & Drug Use, and Sexual Activity domains, CC: Clinical Care, which combines Access to Care and Quality of Care domains. For a full description of each domain, see Appendix Table S1). The ordering of the legend labels is based on alphabetical ordering.

health outcomes. The risk factor indirect or lag effects are translated into how average changes in a risk factor in dependent counties are associated with health in the focal county. In the case of the spatial and network models, dependence is based on geographic adjacency and migrant sending, respectively. In the case of the temporal model, dependence is based on the county's ecological risk factors measured at a prior time period.

Because the primary research objective is to examine diffusion, the tables provide results for the lag effects, with the direct effects provided in Appendix Tables S4-S6. Several key findings emerge. First, county overall health, mortality, and diabetes prevalence are strongly associated with their temporal, spatial and network health lags. For example, a one-unit increase in the mortality rate in a county's prior period, a county's geographically adjacent neighbors, and a county's migration connected communities is associated with a 0.61, 0.31, and 0.55 unit increase in a county's own mortality rate, respectively. When combining geographic distance and migration flows, the association (0.34) falls in between the strictly spatial and network lags.

Second, housing and transit quality, alcohol and drug use, family and social support, and diet and exercise are statistically significant across nearly all specifications. In other words, the lags of these contextual features, whether lagged temporally, spatially or across migrant network pathways, are associated with all health outcomes. Third, the lag effects of the other ecological factors vary across health outcome and clustering type. For example, all health outcomes are associated with temporally lagged smoking and tobacco use, but not with the spatial or network lagged equivalents. Sexual activity is associated with diabetes prevalence across all lag specifications, but has no association with overall health and mortality. Fourth, similar to the correlations presented in Table 1, the coefficient sizes for the combined spatial and migration network model generally fall somewhere in between the coefficients for the strictly spatial and migration network models. However, in many cases, the coefficients are closer in magnitude to the spatial coefficients, indicating a greater influence of spatial proximity relative to migrant connectivity. For example, the combined spatial and migrant overall lag of 0.57 is closer in size to the purely spatial lag (0.46) than the strictly migrant network lag (0.88). This pattern is true for the mortality rate and diabetes prevalence lags.

4. Discussion

Several theoretical frameworks have previously established the clustering of health outcomes and their ecological risk factors (Kolak et al., 2020; Matthews and Yang, 2013; Singer, 1996). A major contribution of this study is to show how these separate yet interrelated frameworks can be expanded by integrating other forms of health clustering and embeddedness. Syndemic theory emphasizes the interaction between multiple health problems, often biologically, with each other and the sociocultural, economic, and physical environment. As Tsai et al. (2017 pp. 1) observed, "although the theory of syndemics is principally a theory about population health, the past two decades' worth of quantitative literature motivated by the theory has generally

Table 1
Levels of temporal, spatial and network clustering of county health outcomes and their risk factors.

	Temporal	Spatial		Network		Spatial + Network	
		Past	Current	Past	Current	Past	Current
Health Factors							
Health Behaviors							
Diet & Exercise	0.83	0.54	0.55	0.41	0.41	0.52	0.53
Tobacco Use	0.58	0.35	0.64	0.27	0.47	0.34	0.61
Alcohol & Drug Use	0.57	0.33	0.51	0.24	0.39	0.32	0.51
Sexual Activity	0.91	0.50	0.46	0.32	0.30	0.43	0.39
Clinical Care							
Access to Clinical Care	0.95	0.20	0.18	0.18	0.16	0.19	0.17
Quality of Clinical Care	0.85	0.51	0.50	0.30	0.29	0.43	0.41
Social and Economic Factors							
Education	0.81	0.42	0.37	0.30	0.24	0.38	0.32
Employment	0.74	0.66	0.61	0.40	0.34	0.57	0.51
Income	0.93	0.55	0.56	0.34	0.34	0.46	0.47
Family and Social Support	0.92	0.52	0.51	0.29	0.28	0.40	0.39
Community Safety	0.85	0.38	0.34	0.25	0.21	0.33	0.28
Physical Environment							
Air & Water Quality	0.14	0.50	0.51	0.36	0.27	0.49	0.42
Housing & Transit	0.91	0.63	0.62	0.37	0.37	0.51	0.52
Health Outcomes							
Overall Health	0.87	0.61	0.66	0.44	0.49	0.57	0.62
Premature Mortality Rate	0.89	0.57	0.55	0.39	0.38	0.51	0.51
Diabetes Prevalence	0.87	0.65	0.65	0.48	0.49	0.61	0.63

Temporal clustering is measured using Pearson's correlation. Spatial and Network clustering are measured using Moran's *I*. Spatial + Network combines migrant and geographic connectivity weight matrices. Past and current variables are measured by the 2015 and 2018 CHR, respectively. The overall health index standardizes and combines the potential life lost before age 75 per 100,000 population, the percentage of adults reporting fair or poor health, average number of physically unhealthy days reported in past 30 days, average number of mentally unhealthy days reported in past 30 days, and percentage of live births with low birthweight. See Appendix Table S1 for years represented in past and current.

Table 2
Transition matrix of county geodemographic classes based on CHR's 13 ecological health risk factors and demographic characteristics.

		Current								
		1	2	3	4	5	6	7	8	N
Past	**1**	**74**	2	4	4	0	2	2	12	560
	2	27	**67**	0	6	0	0	0	0	674
	3	1	0	**92**	6	0	0	0	0	239
	4	3	20	7	**68**	0	1	1	0	738
	5	5	0	0	0	**95**	0	0	0	19
	6	7	27	0	0	0	**63**	3	0	718
	7	1	0	0	16	0	0	**83**	0	130
	8	3	0	0	0	0	0	0	**97**	30

Note: Values represent percent of counties by cluster type across 2015 (past) and 2018 (current) CHR geodemographic classes. Geodemographic classes were constructed using the 13 CHR domains, educational attainment (percent without a high school degree and percent with a college degree), race/ethnicity (percent American Indian, Asian, white, black, and Hispanic), age composition (percent below 18 years old between 25 and 34 years old and 65 years old and over) and percent rural. See Appendix Table S1 for years represented in past and current and Appendix Table S3 for cluster descriptions.

focused on studying individuals rather than populations—and consequently has had very little to say about population health." This study addresses this gap by finding support for syndemic theory at the ecological level. Moreover, whereas prior work has primarily documented the interactions between diseases, this study demonstrates that co-occurrence also applies to the community-level factors theorized to influence health. This finding melds syndemic theory with the SDH framework by demonstrating that co-occurrence is self-reinforcing in that multiple ecological health risk factors may interactively work in concert to diminish or enhance community health. Consequently, targeting a single risk factor may prove to be ineffective because of other determinants collectively impacting local health. However, it is not simply the case that counties can be categorized into those that have health enhancing or health diminishing factors; instead, counties are grouped together based on distinct packages of ecological risk factors, some enhancing and others diminishing. In this case, it is not just independently targeting all 13 of CHR's health factors that is most efficient, but a distinct set of factors depending on the community.

The study's findings also expand the syndemic and traditional SDH frameworks by demonstrating that co-occurrence and the influence of risk factors is not a strictly internal or within community process. The study demonstrates that health outcomes and ecological risk factors also cluster *between* counties. This study highlights the spatial dimensions of co-occurrence, finding high levels of spatial autocorrelation for all health outcomes and risk factors across two time periods. Moreover, the study demonstrates that this clustering has consequences on county health and well-being. This observation is not new, as evidenced by the growing literature demonstrating the spatial dependency of health outcomes (Arcaya et al., 2012; Matthews and Yang, 2013). But this study's findings contributes to this literature by demonstrating that clustering and its influence extend beyond a county's geographic network. Specifically, I find that counties both near and far that are connected via migration flows share similar health factors and outcomes. And a county's health factors and outcomes diffuse across these migration-based pathways. In other words, a county's health is associated with the average levels of health and risk factors in its migration network. This finding is predicted by the established literature on social networks (Christakis and Fowler, 2009). However, rather than relying on individuals as nodes, which is the tradition in this literature, I demonstrate that social network ties have consequences for ecological units and their population-level health.

I also find evidence of strong temporal clustering: current health levels are strongly associated with prior health levels. Moreover, the self-reinforcing structures uncovered across the co-occurrence, spatial and network dimensions also persist over time. A county tends to remain in the same geodemographic classifications over a 5-year period. The magnitude of spatial and network clustering changes little over time. These results indicate that health inequality perseveres partly because communities are embedded in clustered networks that stubbornly endure.

The study does not attempt to rank sources of clustering or answer

Table 3
The association between temporal, spatial and migration network lags and county overall health from Spatial Durbin regression models.

	Temporal	Spatial	Network	Spatial + Network
Overall health lag	0.40***	0.46***	0.88***	0.57**
	(0.01)	(0.02)	(0.03)	(0.02)
CHR Risk Factors lags				
Diet & Exercise	0.10***	−0.11*	−0.84*	−0.13*
	(0.02)	(0.05)	(0.42)	(0.06)
Tobacco Use	0.02**	0.08**	0.15	0.07*
	(0.01)	(0.03)	(0.16)	(0.03)
Alcohol & Drug Use	−0.03**	−0.17***	−0.32	-0.11***
	(0.01)	(0.03)	(0.18)	(0.03)
Sexual Activity	0.01	−0.04	0.17	−0.03
	(0.02)	(0.03)	(0.25)	(0.04)
Access to Clinical Care	0.06	−0.001	−0.04	0.02
	(0.03)	(0.04)	(0.29)	(0.05)
Quality of Clinical Care	0.003	0.03	0.52*	0.11**
	(0.02)	(0.03)	(0.27)	(0.04)
Education	0.02	0.01	−0.23	−0.05
	(0.01)	(0.03)	(0.26)	(0.04)
Unemployment	0.03***	0.01	−0.03	−0.02
	(0.01)	(0.02)	(0.20)	(0.03)
Income	0.01	−0.01	−0.39	−0.03
	(0.02)	(0.03)	(0.26)	(0.04)
Family & Social Support	0.01*	0.02*	0.35*	0.12*
	(0.02)	(0.04)	(0.37)	(0.06)
Community Safety	−0.03	0.10**	0.41	0.12**
	(0.02)	(0.03)	(0.22)	(0.04)
Air & Water Quality	0.01	−0.01	−0.003	0.003
	(0.01)	(0.03)	(0.18)	(0.03)
Housing & Transit	0.04	0.24***	1.59***	0.32***
	(0.03)	(0.05)	(0.50)	(0.06)
AIC	1715.90	1903.50	1743.60	1827.90
BIC	1921.36	2108.92	1949.02	2033.32

Coefficients represent indirect effects estimated from Spatial Durbin regression models. The dependent variable is overall health based on 2018 CHR data. The overall health index standardizes and combines the potential life lost before age 75 per 100,000 population, the percentage of adults reporting fair or poor health, average number of physically unhealthy days reported in past 30 days, average number of mentally unhealthy days reported in past 30 days, and percentage of live births with low birthweight. The overall health lag measures the influence of overall health measured at a prior time period based on 2015 CHR data (temporal), a county's geographically adjacent neighbors (spatial) or a county's migrant sending (network) counties based on 2018 CHR data. Temporal risk factor lags measure the influence of a county's characteristics measured at a previous time period based on 2015 CHR data. Spatial risk factor lags measure the influence of the average characteristics of county's geographically adjacent neighbors based on 2018 CHR data. Network risk factor lags measure the influence of the average characteristics of a county's migrant sending counties based on 2018 CHR data. Spatial + Network combines migrant and geographic connectivity. Standard errors are in parentheses.
***p < 0.001; **p < 0.01; *p < 0.05.

Table 4
The association between temporal, spatial and migration network lags and county premature mortality rates from Spatial Durbin regression models.

	Temporal	Spatial	Network	Spatial + Network
Mortality rate lag	0.61***	0.31***	0.55***	0.34***
	(0.02)	(0.02)	(0.04)	(0.03)
CHR Risk Factors lags				
Diet & Exercise	−0.07*	0.05	0.09	−0.03
	(0.03)	(0.06)	(0.15)	(0.07)
Tobacco Use	−0.07***	0.06	0.04	0.02
	(0.01)	(0.03)	(0.07)	(0.03)
Alcohol & Drug Use	0.04**	0.16***	0.21**	0.12***
	(0.01)	(0.03)	(0.07)	(0.03)
Sexual Activity	−0.04	0.02	−0.14	−0.01
	(0.02)	(0.04)	(0.10)	(0.05)
Access to Clinical Care	0.05	−0.03	0.21	0.004
	(0.05)	(0.05)	(0.11)	(0.05)
Quality of Clinical Care	−0.05*	−0.05	−0.20*	−0.08
	(0.02)	(0.04)	(0.10)	(0.04)
Education	0.01	0.05	0.17	0.09*
	(0.02)	(0.04)	(0.10)	(0.04)
Unemployment	−0.05***	0.03	0.22**	0.11**
	(0.01)	(0.03)	(0.08)	(0.04)
Income	0.06*	−0.12**	−0.15	-0.10*
	(0.02)	(0.04)	(0.10)	(0.04)
Family & Social Support	−0.02*	0.09	−0.05*	−0.01
	(0.03)	(0.05)	(0.14)	(0.05)
Community Safety	0.22***	−0.08*	−0.07	−0.08*
	(0.02)	(0.04)	(0.08)	(0.04)
Air & Water Quality	0.04**	−0.06	−0.05	−0.04
	(0.01)	(0.03)	(0.07)	(0.03)
Housing & Transit	−0.14***	−0.37***	−0.74***	−0.38***
	(0.03)	(0.05)	(0.14)	(0.06)
AIC	3512.50	4493.00	4496.80	4520.30
BIC	3717.97	4698.37	4702.21	4725.70

Coefficients represent indirect effects estimated from Spatial Durbin regression models. The dependent variable is the premature mortality rate based on 2018 CHR data. The mortality rate lag measures the influence of mortality measured at a prior time period based on 2015 CHR data (temporal), a county's geographically adjacent neighbors (spatial) or a county's migrant sending (network) counties based on 2018 CHR data. Temporal risk factor lags measure the influence of a county's characteristics measured at a previous time period based on 2015 CHR data. Spatial risk factor lags measure the influence of the average characteristics of county's geographically adjacent neighbors based on 2018 CHR data. Network risk factor lags measure the influence of the average characteristics of a county's migrant sending counties based on 2018 CHR data. Spatial + Network combines migrant and geographic connectivity. Standard errors are in parentheses.
***p < 0.001; **p < 0.01; *p < 0.05.

whether geography matters more than migration. Rather the study demonstrates that clustering exists in various domains, and either examining this clustering separately or ignoring it completely underestimates a community's disease burden. The study also does not directly speak to health clustering at lower geographic scales such as the neighborhood level. Future research applying this study's analytic framework at the local or neighborhood level can better inform local public health interventions that target micro-level responses.

The study's reliance on multiple data sources is a strength in that it allows for the inclusion of a wide set of ecological health risk factors. However, the CHR data draw from over a dozen sources, which have varying degrees of reliability and vintages. For example, mortality data, which are reported almost 100% of the time, are extremely reliable as counts of death, while other measures (e.g., excessive drinking) are missing for some counties, and still other measures (e.g., air quality and obesity rates) are based on modeling methods. To obtain stable county-level estimates, data from multiple years were combined, meaning that county estimates may not be representative of the current outcome prevalence within a county. Future work using updated data that are relatively consistent across datasets is needed in order to test the robustness of the study's findings. This study is subject to the limitations of the secondary data sets used, including sampling variability and measurement error. However, these data came from large, well-validated surveys, and these limitations have negligible impacts on the study results.

The study cannot make causal claims regarding the pathways connecting the health of a county and the health and ecological risk factors of its spatially adjacent and migration-connected counties. The association between a county's health and the health and ecological risk factors in migration-connected counties may be due to regional economic, legal, political, and historic processes that govern both migration flows and health outcomes. These pathways are not tested in the current study. The study's objective is to descriptively establish clustering at different levels across various domains without speaking to neither the causes of this clustering nor its causal impact on health. Future research that uncover macro-level explanations will help ensure that the conceptual

Table 5
The association between temporal, spatial and migration network lags and county diabetes prevalence from Spatial Durbin regression models.

	Temporal	Spatial	Network	Spatial + Network
Diabetes prevalence lag	0.61***	0.38***	0.65***	0.44***
	(0.02)	(0.02)	(0.03)	(0.02)
CHR Risk Factors lags				
Diet & Exercise	−0.0002	−0.01***	−0.01**	−0.01***
	(0.001)	(0.002)	(0.004)	(0.002)
Tobacco Use	−0.001*	−0.001	−0.0003	−0.002
	(0.0003)	(0.001)	(0.002)	(0.001)
Alcohol & Drug Use	0.001*	0.004***	−0.01**	0.004***
	(0.0003)	(0.001)	(0.002)	(0.001)
Sexual Activity	−0.001*	−0.003*	−0.02***	−0.005***
	(0.001)	(0.001)	(0.003)	(0.001)
Access to Clinical Care	−0.002	−0.003*	−0.003	−0.002
	(0.001)	(0.001)	(0.004)	(0.001)
Quality of Clinical Care	−0.0004	0.004***	0.01***	0.01***
	(0.001)	(0.001)	(0.003)	(0.001)
Education	−0.001	0.001	0.01**	0.004**
	(0.0005)	(0.001)	(0.003)	(0.001)
Unemployment	−0.001	−0.00003	−0.001	0.001
	(0.0003)	(0.001)	(0.002)	(0.001)
Income	0.0003	−0.00003	0.001	−0.001
	(0.001)	(0.001)	(0.003)	(0.001)
Family & Social Support	−0.001*	−0.001*	−0.0005	−0.001
	(0.001)	(0.001)	(0.004)	(0.002)
Community Safety	0.0005	0.001	−0.002	−0.001
	(0.001)	(0.001)	(0.003)	(0.001)
Air & Water Quality	−0.0004	−0.001	0.002	0.001
	(0.0003)	(0.001)	(0.002)	(0.001)
Housing & Transit	−0.002*	−0.01***	−0.02***	−0.01***
	(0.001)	(0.001)	(0.004)	(0.002)
AIC	−19196.20	−18518.21	−18541.30	−18532.11
BIC	−18990.56	−18312.46	−18335.78	−18326.22

Coefficients represent indirect effects estimated from Spatial Durbin regression models. The dependent variable is diabetes prevalence based on 2018 CHR data. The diabetes prevalence lag measures the influence of diabetes prevalence measured at a prior time period based on 2015 CHR data (temporal), a county's geographically adjacent neighbors (spatial) or a county's migrant sending (network) counties based on 2018 CHR data. Temporal risk factor lags measure the influence of a county's characteristics measured at a previous time period based on 2015 CHR data. Spatial risk factor lags measure the influence of the average characteristics of county's geographically adjacent neighbors based on 2018 CHR data. Network risk factor lags measure the influence of the average characteristics of a county's migrant sending counties based on 2018 CHR data. Spatial + Network combines migrant and geographic connectivity. Standard errors are in parentheses.
***$p < 0.001$; **$p < 0.01$; *$p < 0.05$.

framework presented in this study can be used to address underlying causes.

The study focuses on one type of network ecological tie – migration – but other phenomena tying communities together may also exhibit clustering including work commuting, activity spaces, and economic development (Cagney et al., 2020), and these various linkages may have different consequences on a community's health. Future work examining the clustering and consequences of these different networks will help expand and generalize this study's framework. The study's findings also do not apply to counties in Alaska and Hawaii as they were excluded from the analyses due to methodological reasons related to the construction of the spatial and migration network matrices.

Despite these limitations, the study's findings as they are substantiated with more recent data and at multiple scales, could help to inform national, state and county public health responses that look at groups of health determinants and outcomes concurrently. The overarching framework of this research is that ecological health and its risk factors are dependent on multiple forms of embeddedness, and an understanding of the various dimensions by which health is clustered represents a tool for explaining patterns of persistent poor health and developing novel approaches to addressing health inequalities. The study offers potential ways to expand current approaches that rely largely on a single theoretical framework or method of intervention. A syndemic-oriented intervention defines the population in question, identifies the conditions that create and sustain health in that population, examines why those conditions might differ among groups and determines how those conditions might be addressed in a comprehensive manner. For example, an intervention targeting chronic alcoholism/HIV infection syndemic includes the simultaneous targeting of drinking patterns among individuals living with HIV infection, focusing on drinking establishments as sites of prevention education, the close monitoring of dually diagnosed individuals to ensure appropriate treatment, and the provision of training that assists individuals in handling cognitive and motor impairments (Singer et al., 2012). An SDH approach focuses on the ecological context potentially influencing poor health outcomes, such as housing, school characteristics, aspects of the built environment and crime and safety. The syndemic and SDH approaches focus on the co-occurrence of health and its risk factors, but this study's findings illuminate other dimensions through which health outcomes and their risk factors are clustered. Specifically, a syndemic or SDH based intervention in a community may incorporate that community's larger spatial and network community, which can be done numerous ways. For example, spatial and network clusters can provide appropriate comparison groups that health agencies can use for benchmarking as a forum for learning exchange. Local health departments often look to metrics such as the CHR as a means to evaluate their performance nationally or in relation to their same-state peers. However, comparisons of different populations across the nation or within a state may obscure net gains being made by health departments serving more disadvantaged populations. The framework established in this study provides a mechanism to create comparison groups, through which the lowest performers within a cluster can gain valuable information on how to improve their performance by implementing similar interventions as the highest performers in the cluster. Embedded counties can also collaborate on interventions, sharing resources to mitigate risk factors influencing health in their clustered communities.

Social network based interventions traditionally focus on individual social ties. For example, obesity interventions use network ties established via online networks or mobile apps to encourage physical activity and promote other health behaviors associated with proper weight loss (Jane et al., 2018). This study expands the traditional social network approach by demonstrating the importance of networks at the community level, exposure to health promoting or impeding factors in spatially embedded communities, and the interaction between multiple health outcomes and risk factors that may spread across a community network. That is, it is not just the network of individuals and their characteristics that a person is exposed to that may increase or minimize unhealthy behaviors, but also the network of communities and their ecological risk factors that a person moves through, whether via residential migration or other forms of spatial mobility.

Spatial diffusion based interventions focus on selectively targeting hotspot areas. That is, similar to crime interventions that, for example, distribute policing to specific geographic clusters exhibiting high crime rates, public health interventions will target geographically contiguous areas showing significant poor health. The rationale for this approach is that state and local public health officials can focus limited resources in areas with concentrated health disadvantages. This study expands this approach by demonstrating that clustering between communities can also occur across networks connecting distal counties. In this case, practitioners from both geographically adjacent and network connected communities can collaborate to share and deploy resources across their respective communities. Syndemic, SDH, social network and spatial diffusion approaches often neglect temporal aspects of health clustering. Measuring temporal clustering provides agencies predictive models to understand the health, social and economic trajectories of their populations, and to use these models to develop strategies that break the

chains that bind health problems across time.

How to best address rising health inequality is a complex question given the local social and economic context and the multiple strategies that can be implemented. Much of the discussion surrounding the causes and consequences of increasing mortality and decreasing life expectancy in the United States has focused on issues related to medical care and individual behaviors and choices. Although individual factors are an important part of the story, the macro-level health trends require macro-level explanations (Zajacova and Montez, 2017). This study's results suggest that scholars and practitioners should move beyond the evaluation and targeting of single outcomes, determinants and communities. Health and its ecological risk factors are clustered phenomena, and in order to fully address population health inequality, we must break down the self-reinforcing clustered networks that allow it to stubbornly persist.

Credit author

Noli Brazil: Conceptualization, Methodology, Software, Data curation, Writing- Original draft preparation, Visualization, Investigation, Writing- Reviewing and Editing.

Acknowledgements

I thank Catherine Brinkley and participants at the 2019 Population Association of American conference for their valuable feedback on earlier drafts of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.socscimed.2021.113772.

References

Anselin, L., 1988. Spatial Econometrics: Methods and Models, vol. 4. Springer Science & Business Media.
Arcaya, M., Brewster, M., Zigler, C.M., Subramanian, S.V., 2012. Area variations in health: a spatial multilevel modeling approach. Health Place 18 (4), 824–831.
Bambra, C., Gibson, M., Sowden, A., Wright, K., Whitehead, M., Petticrew, M., 2010. Tackling the wider social determinants of health and health inequalities: evidence from systematic reviews. J. Epidemiol. Community Health 64 (4), 284–291.
Barker, L.E., Thompson, T.J., Kirtland, K.A., Boyle, J.P., Geiss, L.S., McCauley, M.M., Albright, A.L., 2013. Bayesian small area estimates of diabetes incidence by United States county, 2009. J. Data Sci.: JDS 11 (1), 269.
Barnes, P.M., Adams, P.F., Powell-Griner, E., 2010. Health Characteristics of the American Indian or Alaska Native Adult Population: United States, 2004-2008, vol. 20. National Health Statistics Report, pp. 1–22.
Bastomski, S., Brazil, N., Papachristos, A.V., 2017. Neighborhood co-offending networks, structural embeddedness, and violent crime in chicago. Soc. Network. 51, 23–39.
Bojanowski, M., Corten, R., 2014. Measuring segregation in social networks. Soc. Network. 39, 14–32.
Brazil, N., 2017. Spatial variation in the Hispanic Paradox: mortality rates in new and established Hispanic US destinations. Popul. Space Place 23 (1), e1968.
Brazil, N., Clark, W.A., 2017. Individual mental health, life course events and dynamic neighbourhood change during the transition to adulthood. Health Place 45, 99–109.
Cagney, K.A., Cornwell, E.Y., Goldman, A.W., Cai, L., 2020. Urban mobility and activity space. Annu. Rev. Sociol. 46.
Charrad, M., Ghazzali, N., Boiteau, V., Niknafs, A., Charrad, M.M., 2014. Package 'nbclust'. J. Stat. Software 61, 1–36.
Christakis, N.A., Fowler, J.H., 2007. The spread of obesity in a large social network over 32 years. N. Engl. J. Med. 357 (4), 370–379.
Christakis, N.A., Fowler, J.H., 2009. Connected: the Surprising Power of Our Social Networks and How They Shape Our Lives. Little, Brown Spark.
Clark, W.A., Brazil, N., 2019. Neighborhood selections by young adults: evidence from a panel of US adolescents. J. Urban Aff. 41 (7), 981–998.
Clark, W.A., Maas, R., 2015. Interpreting migration through the prism of reasons for moves. Popul. Space Place 21 (1), 54–67.
Danchev, V., Porter, M.A., 2018. Neither global nor local: heterogeneous connectivity in spatial network structures of world migration. Soc. Network. 53, 4–19.
Delmelle, E.C., 2019. The increasing sociospatial fragmentation of urban America. Urban Science 3 (1), 9.
DeWaard, J., Ha, J.T., 2019. Resituating relaunched migration systems as emergent entities manifested in geographic structures. Migration Studies 7 (1), 39–58.
Dwyer-Lindgren, L., Bertozzi-Villa, A., Stubbs, R.W., Morozoff, C., Mackenbach, J.P., van Lenthe, F.J., et al., 2017. Inequalities in life expectancy among US counties, 1980 to 2014: temporal trends and key drivers. JAMA Internal Medicine 177 (7), 1003–1011.
Fawcett, J.T., 1989. Networks, linkages, and migration systems. Int. Migrat. Rev. 23 (3), 671–680.
Fenelon, A., 2013. Geographic divergence in mortality in the United States. Popul. Dev. Rev. 39 (4), 611–634.
Graif, C., Freelin, B.N., Kuo, Y.H., Wang, H., Li, Z., Kifer, D., 2019. Network spillovers and neighborhood crime: a computational statistics analysis of employment-based networks of neighborhoods. Justice Q. JQ 1–31.
Grubesic, T.H., Miller, J.A., Murray, A.T., 2014. Geospatial and geodemographic insights for diabetes in the United States. Appl. Geogr. 55, 117–126.
Gruebner, O, Lowe, S, Sampson, L, Galea, S, 2015. The geography of post-disaster mental health: spatial patterning of psychological vulnerability and resilience factors in New York City after Hurricane Sandy. Int. J. Health Geogr. 14 (1).
Hood, C.M., Gennuso, K.P., Swain, G.R., Catlin, B.B., 2016. County health rankings: relationships between determinant factors and health outcomes. Am. J. Prev. Med. 50 (2), 129–135.
Jane, M., Hagger, M., Foster, J., Ho, S., Pal, S., 2018. Social media for health promotion and weight management: a critical debate. BMC Publ. Health 18 (1), 1–7.
Kolak, M., Bhatt, J., Park, Y.H., Padrón, N.A., Molefe, A., 2020. Quantification of neighborhood-level social determinants of health in the continental United States. JAMA network open 3 (1) e1919928-e1919928.
Kramer, M, Black, N, Matthews, S, James, S, 2017. The legacy of slavery and contemporary declines in heart disease mortality in the US South. SSM-Popul. Health 3, 609–617.
Leal, D.F., Malhotra, R.S., Misra, J., 2019. Visualizing feminized international migration flows in the 1990s. Socius 5, 2378023118819940.
Lee, Y., Ogburn, E.L., 2020. January). Testing for network and spatial autocorrelation. International Conference on Network Science. Springer, Cham, pp. 91–104.
LeSage, J., Pace, R.K., 2009. Introduction to Spatial Econometrics. Chapman and Hall.
Lichter, D.T., Johnson, K.M., 2009. Immigrant gateways and Hispanic migration to new destinations. Int. Migrat. Rev. 43 (3), 496–518.
Markides, K.S., Coreil, J., 1986. The health of Hispanics in the southwestern United States: an epidemiologic paradox. Publ. Health Rep. 101 (3), 253.
Marmot, M., 2005. Social determinants of health inequalities. Lancet 365 (9464), 1099–1104.
Matthews, S.A., Yang, T.C., 2013. Spatial polygamy and contextual exposures (spaces) promoting activity space approaches in research on place and health. Am. Behav. Sci. 57 (8), 1057–1081.
Meader, N., King, K., Moe-Byrne, T., Wright, K., Graham, H., Petticrew, M., et al., 2016. A systematic review on the clustering and co-occurrence of multiple risk behaviours. BMC Publ. Health 16 (1), 657.
Mendenhall, E., 2016. Beyond comorbidity: a critical perspective of syndemic depression and diabetes in cross-cultural contexts. Med. Anthropol. Q. 30 (4), 462–478.
Menke, A., Casagrande, S., Geiss, L., Cowie, C.C., 2015. Prevalence of and trends in diabetes among adults in the United States, 1988-2012. Jama 314 (10), 1021–1029.
Myers, C, Slack, T, Broyles, S, Heymsfield, S, Church, T, Martin, C, 2017. Diabetes prevalence is associated with different community factors in the diabetes belt versus the rest of the United States. Obesity 25 (2), 452–459.
Norris, T., Vines, P.L., Hoeffel, E.M., 2012. The American Indian and Alaska Native Population: 2010. US Department of Commerce, Economics and Statistics Administration, US Census Bureau, Washington, DC, p. 21.
Papachristos, A.V., Bastomski, S., 2018. Connected in crime: the enduring effect of neighborhood networks on the spatial patterning of violence. Am. J. Sociol. 124 (2), 517–568.
Remington, P.L., Catlin, B.B., Gennuso, K.P., 2015. The county health rankings: rationale and methods. Popul. Health Metrics 13 (1), 11.
Sampson, R.J., 2012. Great American City: Chicago and the Enduring Neighborhood Effect. University of Chicago Press.
Sharkey, P., 2008. The intergenerational transmission of context. Am. J. Sociol. 113 (4), 931–969.
Singer, M., 1996. A dose of drugs, a touch of violence, a case of AIDS: conceptualizing the SAVA syndemic. Free Inq. Creativ. Sociol. 24 (2), 99–110.
Singer, M., Bulled, N., Ostrach, B., 2012. Syndemics and human health: implications for prevention and intervention. Ann. Anthropol. Pract. 36, 205–211, 2012.
Singer, M., Bulled, N., Ostrach, B., Mendenhall, E., 2017. Syndemics and the biosocial conception of health. Lancet 389, 941–950, 10072.
Singleton, A.D., Spielman, S.E., 2014. The past, present, and future of geodemographic research in the United States and United Kingdom. Prof. Geogr. 66 (4), 558–567.
Sparks, P, Sparks, C, 2010. An application of spatially autoregressive models to the study of US county mortality rates. Popul. Space Place 16 (6), 465–481.
Tabb, L., McClure, L., Quick, H., Purtle, J., Roux, A., 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.
Tam, C.C., 2019. Beyond the immigrant enclave: differentiating between coethnic neighborhoods for the study of health and social problems. Geojournal 84 (4), 983–999.
Tita, G.E., Radil, S.M., 2010. Spatial regression models in criminology: modeling social processes in the spatial weights matrix. In: Handbook of Quantitative Criminology. Springer, New York, NY, pp. 101–121.
Tsai, A.C., 2018. Syndemics: a theory in search of data or data in search of a theory? Soc. Sci. Med. 206, 117–122.

Tsai, A.C., Venkataramani, A.S., 2016. Syndemics and health disparities: a methodological note. AIDS Behav. 20 (2), 423–430.
Tsai, A.C., Mendenhall, E., Trostle, J.A., Kawachi, I., 2017. Co-occurring epidemics, syndemics, and population health. Lancet 389, 978–982, 10072.
Umberson, D., Karas Montez, J., 2010. Social relationships and health: a flashpoint for health policy. J. Health Soc. Behav. 51 (1_Suppl. l), S54–S66.
Vierboom, Y.C., Preston, S.H., Hendi, A.S., 2019. Rising geographic inequality in mortality in the United States. SSM-Population Health 9, 100478.
Wallace, M., Sharfstein, J.M., Kaminsky, J., Lessler, J., 2019. Comparison of US county-level public health performance rankings with county cluster and national rankings: assessment based on prevalence rates of smoking and obesity and motor vehicle crash death rates. JAMA network open 2 (1) e186816-e186816.
Wickrama, K.A., Conger, R.D., Wallace, L.E., Elder Jr., G.H., 1999. The intergenerational transmission of health-risk behaviors: adolescent lifestyles and gender moderating effects. J. Health Soc. Behav. 40 (3), 258.
Williams, D.R., Sternthal, M., 2010. Understanding racial-ethnic disparities in health: sociological contributions. J. Health Soc. Behav. 51 (1_Suppl. l), S15–S27.
Xu, J., Murphy, S.L., Kochanek, K.D., Arias, E., 2016. Mortality in the United States, 2015. National Center for Health Statistics, Hyattsville, MD.
Yang, T.C., 2019. Modeling the importance of within-and between-county effects in an ecological study of the association between social capital and mental distress. Prev. Chronic Dis. 16.
Yang, T.C., Noah, A.J., Shoff, C., 2015. Exploring geographic variation in US mortality rates using a spatial Durbin approach. Popul. Space Place 21 (1), 18–37.
Zajacova, A., Montez, J.K., 2017. Macro-level perspective to reverse recent mortality increases. Lancet 389, 991–992, 10073.