WORKING
PAPER

Neighborhood Archetypes for Population Health Research

Is There No Place Like Home?

MARGARET M. WEDEN
CHLOE E. BIRD
JOSE J. ESCARCE
NICOLE LURIE

WR-738

January 2010

This paper series made possible by the NIA funded RAND Center for the Study of Aging (P30AG012815) and the NICHD funded RAND Population Research Center (R24HD050906).

RAND LABOR AND POPULATION

Neighborhood Archetypes for Population Health Research: Is There No Place Like Home?

Margaret M. Weden, PhD[1]; Chloe E. Bird, PhD[1]; José J Escarce, MD, PhD[1,2]; Nicole Lurie, MD, MSPH[3]

[1]RAND Corporation; [2]University of California, Los Angeles; [3]U.S. Department of Health and Human Services. We acknowledge the contributions of Ricardo Basurto-Davila and Adria Dobkin for helping prepare the data; Paul Steinberg for editing; and Richard Carpiano, Stephanie Robert, and Erin Ruel for helpful suggestions on earlier versions of the manuscript. This work was supported by a grant from the National Institute on Environmental Health Sciences (P50 ES012383).

ABSTRACT

The principal objective of this study is to characterize the places in which people live by factors associated with physical and mental wellbeing. We demonstrate a new approach that employs neighborhood measures such as social environment, built environment, commuting and migration, and demographics and household composition to classify neighborhoods into archetypes. The number of neighborhood archetypes, their defining attributes, and their change/stability between 1990 and 2000 is analyzed using latent class analysis applied to a rich array of data sources. In both years, six archetypes of U.S. neighborhoods are differentiated which occur at prevalence from 13% to 20%: Mobile single-household, urbanites; Low SES, rural; Poor, urban, minority; Low SES, urban, minority commuters; High SES, foreign born, new home owners; and Middle-class suburban/exurban families. Findings show that neighborhoods have remained notably constant between 1990 and 2000, with 76.4% of the neighborhoods categorized as the same archetype ten years later. The approach to defining neighborhood archetypes translates the theoretical aspects of research on neighborhoods and health into a measurement typology that can be employed in applied research questions such as public health surveillance and planning and which can be replicated and extended for use in other historical, geographical, and substantive applications.

INTRODUCTION

Research on neighborhoods and health is motivated by the idea that we live in places that represent more than physical locations. They are also the manifestation of the social, cultural, political and geographic cleavages that shape a constellation of risks and resources. Research on neighborhood effects has reconnected public health with its earlier population foundations—showing that the social ecology and built environments are important "upstream" determinants of chronic and infectious disease. This work documents how social and built environments structure opportunities and barriers to more proximal social and material determinants of health (Sampson et al. 2002; Cummins et al. 2007).

Neighborhoods and health research draws heavily on theory and methodologies from Chicago School factorial social ecology which relied on Census data and measures and which developed factor analysis (e.g. Janson 1980; Schwirian 1983; for critique see Sampson et al. 2002). This approach conceptualized four primary axes of neighborhood structure—class, race/ethnicity, density, and life-course stage. The theory and methods also informed the most commonly employed measures of neighborhoods for neighborhoods and health research (Sampson et al. 2002).

In this paper, we reconsider the models and measures of neighborhoods that emerged from the Chicago School factorial social ecology and explore whether there have been changes in the social ecology of neighborhoods since the four primary cleavages were identified. We address questions raised in literature reviews on the characteristics of US neighborhoods, the relevance of the built environment, and the dynamics of neighborhoods over time (Diez Roux 2001; Sampson et al. 2002; Robert et al. forthcoming). We develop a new, complementary theoretical and methodological approach to study neighborhoods that employs latent class analysis to characterize neighborhood archetypes and assess stability or change. In so doing, we produce a reliable measure of U.S. neighborhood archetypes that can be employed in future research on neighborhoods and health.

Neighborhood Social Ecology and Constructs

Researchers have identified how economic, social, demographic, geographic, structural, and institutional conditions of a neighborhood coalesce to influence physical and mental wellbeing. While some studies highlight specific neighborhood characteristics—e.g. neighborhood poverty (Haan et al. 1987), racial and ethnic concentration (Collins & Williams 1999), or urbanization (Galea & Vlahov 2005)—most indicate that multiple factors affect neighborhood characteristics. Multidimensional constructs have been used to examine how socioeconomic affluence or disadvantage, and social disorganization impact individual's health and social wellbeing (e.g., Sampson et al. 2002; Cummins et al. 2007). Examples of indicators include neighborhood socioeconomic status (NSES), (concentrated) neighborhood affluence and/or neighborhood disadvantage, and neighborhood social cohesion, collective efficacy and social disorganization (Sampson et al. 2002).

Neighborhood indicators employed in population health research typically measure social context using poverty rates, percentage of residents receiving public assistance, percentage of female-headed families, unemployment ratio, and percentage of African American residents (for review see Browning & Cagney 2002). Studies have also extended the NSES construct to describe how neighborhood disadvantage and neighborhood affluence exert independent effects on individual health highlighting the relevance of high income, education and occupational status in so doing (e.g., Weden et al. 2008).

Similarly, work on structural aspects of neighborhoods and the relevance of social disorganization (e.g. as indicated by boarded-buildings, vacancy rates, and residential turnover; Wilson & Kelling 1982) extends a long history of research (beginning with seminal works by Durkheim and Simmel) that relates rural-urban differentials and industrialization to social disorganization, and therein, physical and mental health (see review Vlahov & Galea 2002). Neighborhood residential turnover also has been linked to poor child development, problem behavior, and health-related risks (for review see Jelleyman & Spencer 2008).

Additionally, urban planning, urban studies and social ecology provide a theoretical foundation linking social and physical dimensions of the neighborhood (e.g. see reviews by Vlahov 2002; Corburn 2004). Research on the neighborhood life cycle links shifts in the demographic composition of communities to changing land-use patterns (i.e. from residential to commercial; see Downs 1981). Sociological research links residential turnover and deterioration of physical infrastructure to social disorganization (e.g. Sampson and Groves 1989).

Recent studies refocused attention to built environment factors that support active life styles and reduce the risk of chronic disease, such as land use, commuting patterns and walkability (see reviews by Frumkin 2003; Srinivasan et al. 2003; Galea et al. 2005). Yet few studies have reconsidered the linkages between neighborhood social ecology and built environment in light of current population health dynamics in chronic disease (see review Diez Roux 2001; Galea et al. 2005). One notable exception is research on social capital and the built environment as it relates to physical activity and obesity (e.g. Leyden 2003; Poortinga 2006; Wood & Giles-Corti 2008; Cohen et al. 2008).

Theoretical Motivation for Neighborhood Archetypes

The first wave of studies on neighborhoods and health focused on showing that 'neighborhoods matter' and have independent effects beyond individual socioeconomic characteristics. These studies argued that neighborhoods influence health and behavior through mechanisms such as collective socialization, peer-group influence, and institutional capacity. The second wave of studies on neighborhoods and health evaluated these mechanisms with latent measures of neighborhood characteristics (such as level of segregation, collective social and economic capacity, or social disorganization) (Sampson et al. 2002). In this work, factor analysis or structural equation models are used to create scales for these characteristics and identify a continuum of sociodemographic disadvantage or affluence on which neighborhoods were located. We call this approach a 'variable perspective' to neighborhood research.

Although the variable perspective is useful for answering questions about the independent effect of specific neighborhood characteristics controlling for individual characteristics, it is not as well suited to studying how various aspects of neighborhoods combine to effect health and whether and how the effects differ over the life course. Rather than being defined by a single dimension, neighborhoods are the synthesis of different combinations of social, economic, demographic, structural and geographic conditions, which affect individuals' lives and health. For example, the impact of neighborhood poverty depends on the community's level of urbanization, age composition, and degree of segregation (Jargowsky 1997; Boardman et al. 2005). Similarly neighborhood socioeconomic disadvantage is associated with and can be exacerbated by environmental risk factors including pollution and environmental hazards (Cutter & Scott 2000; Ponce et al. 2005).

To date, most work has employed this 'variable perspective'. For example, Boardman and colleagues (2005) model the interacting effects of poverty and racial segregation, addressing two of the four axes of neighborhood in the foundational work on neighborhood social ecology (Janson 1980; Schwirian 1983). The problem of the variable approach become evident when additional axes are considered (e.g., a simple model with single dichotomous indicators for each of the four axes requires 4 main effects and 12 interactions), and even more so when additional dimensions of the neighborhood (e.g. the built environment) are considered. Adequate statistical power for interactions between all of these neighborhood variables quickly becomes unattainable. The analytical problem of the 'variable perspective' to neighborhood research is analogous to the 'variable problem' in life course research described previously by Singer and colleagues (1998).

Rather than producing neighborhood measures that decompose residential environments into their parts, we propose identifying archetypal neighborhoods that describe this interacting synthesis of components.

Addressing Outstanding Questions about Neighborhood Classification with LCA

There are two areas of research on neighborhoods and health our approach is well designed to extend. The first pertains to the interactions between different conceptual dimensions of the neighborhood. Interactions between conceptual dimensions can be studied in social science by characterizing archetypes. The empirical method of latent class analysis (LCA) was designed for characterizing archetypes (e.g., Hagenaars & Halman 1989) and has been used extensively in social, behavioral, and health research (Bollen 2002). The second area of research pertains to temporal dynamics including neighborhood change (e.g. gentrification, racial succession) and neighborhood life cycles (e.g., Schwirian 1983; Sampson 2002; Robert et al. forthcoming).

To date, LCA has not been applied to neighborhood characterization for population health research, yet the approach offers distinct analytic advantages to alternative methods previously employed (e.g., factor analytic methods, including structural equation modeling (SEM), and cluster analysis techniques). The advantages of LCA are reviewed elsewhere (Rapkin et al. 1993; Chow 1998), and pertain directly to two new areas of research on neighborhoods and health. First, LCA can measure how constellations of characteristics capture distinct neighborhood archetypes. These constellations of characteristics are described by 'interactions' between neighborhood dimensions. Thus LCA allows a researcher to identify the most statistically robust set of interactions between dimensions as a constellation of characteristics that describe the places of interest. Secondly, like factor analytic methods (e.g., factor analysis, SEM), LCA allows one to assess the stability or change of neighborhood archetypes independent of the measurement of the neighborhood archetypes at a given point in time (and can capture change over time). In contrast with these methods, though, LCA can not only be used to statistically test whether the distribution of neighborhood archetypes in a population changes over time, but it can be used to characterize neighborhoods discretely into archetypes at different points in time. Once individual neighborhoods are discretely characterized, questions about the life cycles of individual neighborhoods can be explored. In summary, at its minimum, LCA is a data reduction mechanism similar to cluster analysis. In its full application, LCA becomes a powerful tool for the characterization of neighborhood archetypes and analysis of neighborhood change.

METHODS

Data and Sample

Data on U.S. neighborhoods come from a neighborhood characteristics database compiled and disseminated by RAND Corporation (http://www.rand.org/health/centers/pophealth/data.html). The neighborhood characteristics database contains a rich array of contextual data from the 1990 and 2000 Decennial Census, the Census Topologically Integrated Geographic Encoding and Referencing (TIGER/Line) files, the Environmental Protection Agency Air Quality System, and the American Chamber of Commerce Research that has been compiled, harmonized, and documented by RAND. U.S. neighborhoods are defined at the geographical level of the census tract, with harmonization for changes in tract definitions between 1990 and 2000. Models are estimated using 20% random samples of the complete set of U.S. census tracts in each year, so that 12,252 tracts are observed in 1990 and 13,261 tracts are observed in 2000.

Neighborhood Characteristics by Domain

Indicators of the neighborhood characteristics are selected that: (1) are theoretically related to population health; (2) entail previously validated variables of the social and built environment; and (3) were measured identically in 1990 and 2000. Specific variables fall into four domains: built environment, migration and commuting, socioeconomic composition, and demographics and household composition.

Table 1 details how categorical indicators were constructed for each neighborhood measure described below. Refinement of the indicators was conducted in conjunction with refinement of the measurement models described below and in Appendix 2. This included sensitivity analyses to evaluate the improvement of models achieved by recategorizing continuous variables as categorical or dichotomous indicators, combining indicators, and dropping redundant indicators (see Appendix 2).

Built Environment

Urbanization is measured using *density* (population per square kilometer) and *urbanicity* (or % rural) categorized as exclusively rural, exclusively urban (100% versus 0% rural respectively), or mixed (suburban, exurban or urbanizing). Land-use patterns are measured via mean *block size*, the number of intersections or "*nodes*," and two measures of *walkability* (the gamma and alpha index of street connectivity; Taaffe and Gauthier 1973). The quality and upkeep of neighborhood infrastructure is measured via the *mean value of the housing stock*, *percent owner-occupied dwellings*, the *mean housing construction date,* and *percent vacant dwellings*. *Air quality* (expressed by the concentration of particulate matter smaller than 10 micrometers, or PM10) is included as in indicator of environmental pollution, using a threshold for health compromising levels of 50 ug/m^3 PM10 (from yearly averages) based on previous findings (Daniels et al. 2000).

Migration and Commuting

Indicators of internal and external migratory patterns capture the relevance of instability in the home environment (Wyly 1999) and are measured through *residency* and *housing turnover*. Commuting patterns are a new dimension of the neighborhood in population health research that has been related to opportunities for physical activity and exposure to psychosocial stress (for review see Hamer & Chida 2008). The indicators of commuting are divided according the *length of commute* and are complemented by the *mode of transportation to work*.

Socioeconomic Composition

Socioeconomic composition captures an axis of the social ecology model that has been more recently validated in measures of NSES, neighborhood affluence, and neighborhood disadvantage, as described earlier. The indicators entail *educational attainment*, *labor force characteristics*, and *economic characteristics*.

Demographic and Household Composition

The demographic and household composition domain captures the last two domains of the social ecology model—racial/ethnic composition and life-cycle stage. The demographic composition is described by *race and ethnicity, native language* and *age*. Household composition refers to household structure and is captured through the *proportion of singles*, *large families*, and *female-headed households*.

Latent Class Model

LCA models are used to identify, characterize, and measure the latent, unobserved categorical variable for the neighborhood archetypes. Neighborhood archetypes are modeled separately for 1990 and 2000, and then a multigroup LCA model is fit to the data from both years combined. The multigroup LCA model can be used to assess whether the distribution of neighborhood archetypes and their characterization changes over time. On the basis of findings from the LCA models for 1990 and 2000, we hold characterization of the neighborhood archetypes constant in the final multigroup LCA model, and we test whether the distribution of neighborhoods across neighborhood archetypes changes between 1990 and 2000.

The LCA models are fit using observed data on the indicators of the built environment, migration and commuting, socioeconomic composition, and demographic and household composition described earlier (see also Table 1). The data are used to characterize the unobserved latent variable for the neighborhood archetypes. Refinement of the structural component of the LCA models (e.g. the number of neighborhood archetypes) and the measurement components of the LCA models (e.g. the characteristics of the neighborhood archetypes) are considered iteratively until the best fitting LCA model is identified (Hagenaars & McCutcheon 2002). Goodness of fit statistics (e.g. the Lo-Mendell-Rubin likelihood ratio test, the Bayesian Information Criteria, and entropy measures) and statistical tests of significance for model parameters are used in the refinement of the structural and measurement models (Hagenaars 1990 Ramaswamy et al. 1993; Lo et al. 2001).

Mplus software version 4.2 (Mplus Version 4.2. 2006) is used to fit the LCA models, accounting for missing variables on some observations. It is also employed to predict latent class membership using the findings from the final LCA multigroup model. These findings allow us to produce a dataset in which every census tract in the U.S. has been probabilistically assigned to the best fitting neighborhood archetype.

Detail on the LCA modeling approach is described in the appendices, specifically, the statistical details of the LCA models (Appendix 1), refinement of the LCA models and sensitivity analyses (Appendix 2).

FINDINGS

How many neighborhood archetypes are there in 1990 and 2000?

Six neighborhood archetypes best summarize the combinations of patterns of neighborhood characteristics data, and this number of archetypes does not change from 1990 to 2000. The findings on the number of neighborhood archetypes for the LCA models fit for each year in Table 2 show that a six-class model produces the best goodness of fit statistics for the respective models for each year. The Lo-Mendell-Rubin test demonstrates that the improvement in the fit of each of these LCA models is statistically significant when contrasting the five- and six-class models, but not when contrasting the fit of the six- and seven-class models.

It is important to note that these goodness-of-fit statistics on the number of neighborhood archetypes (i.e. the 'structural component' of the model) are detailed for the final models for each year identified after iteratively improving the structural and measurement components of the LCA models as described in the methods section. Furthermore, sensitivity analyses (in Appendix 2) validate that six neighborhood archetypes produce the best fitting model in each year.

What are the defining characteristics of the neighborhood archetypes?

A summary of the characteristics for all of the neighborhood archetypes is provided in Table 3. It condenses the detailed multigroup LCA model findings for all parameters of the measurement model in Appendix 3. In Table 3, characteristics are listed for which the probability of the characteristic is greater than 0.6 in the class, as reported in Appendix 3. Characteristics are organized by domain and reported by class. Within each class and domain, they are reported in descending order of probability. Finally, characteristics are denoted with bold archetype where the probability of the characteristic is highest across the classes. For example, Table 3 lists three characteristics in the built environment domain have a probability above 0.6 for neighborhood type 1: urbanicity, density, and walkability. Given type 1 membership, the probability of being categorized as urban is 0.921, the probability of density above the national median is 0.841, and the probability of walkability above the national median is 0.729 (Appendix 1, Column 1). Urban is listed first because it is the most likely of the built environment characteristics for type 1, and it is denoted in bold because the probability of urban is highest across all the classes. Notice that the likelihood of urban is also 0.921 for neighborhood type 4, so urban is also bolded in the list of build environment characteristics for type 4.

As follows, we describe the findings of the most prevalent neighborhood characteristics summarized from Appendix 3 in Table 3. We also describe the least common characteristics, that are detailed in Appendix 3 but not included in the Table 3 summary.

Neighborhood Type 1: Mobile Single-household, Urbanites

The first neighborhood archetype has high density and urbanicity and is highly walkable, as noted above. These neighborhoods are also defined by low levels of vacancy and relatively old housing stock. They have the highest likelihood of high turnover and low residency (i.e. unstable migrants). The socioeconomic status of these communities is high, with both a large probability of 'mid-high' educational composition and a small probability of poverty, public assistance, and unemployment. Among all the neighborhoods, this is the least likely to have a high prevalence of children (age 0–17 years) but the most likely to contain elderly adults (age 80 and older) and young adults (age 18–34 years). Overall, these neighborhoods are most predominantly characterized by having the highest likelihood of one-person households.

Neighborhood Type 2: Low SES Rural

The second neighborhood archetype is the most likely to be rural. It is also the most likely to have vacancy levels above the national median and to have residents working at home. The socioeconomic status is low, as it is the archetype most likely to have ”mid-low” education and it also has likely to have 'mid-low' income and high levels of poverty, public assistance, and males out of the labor force. The demographic composition of these neighborhoods is stratified with both a high proportion of young persons and of older persons. It has he highest likelihood of above median proportion of adults in the early retirement years (e.g., age 65–80). Neighborhoods of this archetype are predominantly white and the least likely to have a large black population.

Neighborhood Type 3: Poor, Urban Minority

The third neighborhood archetype is moderately dense and urban. These neighborhoods are the most likely to have walkable streets, commuting times are typically short, and there is a high likelihood in these neighborhoods that working residents commute by walking or biking. However, these neighborhoods are also the most likely to have poor housing stock in terms of value, age (or the date of construction), and levels of vacancy. They are the most likely to have high housing turnover coupled with low rates of out-migration (i.e. unstable residents). While walkability and commuting are good, joblessness is high (in terms of men out of the labor market and unemployment). On nearly all socioeconomic indicators, these neighborhoods are below the national average—rates of unemployment, public assistance, and poverty are all above the national median, and the income composition is the most likely among all archetypes to be "mid-low." Sociodemographically, these neighborhoods have a high concentration of black and Spanish-speaking residents, and they are the most likely to have female-headed households.

Neighborhood Type 4: Low SES, Urban Minority Commuters

The fourth neighborhood archetype was among the two most urban and dense archetypes. This neighborhood archetype is also most likely to have hazardous air quality, although the overall likelihood of hazardous PM10 levels across all six of the neighborhood archetypes is quite low. Residents of these neighborhoods are the most likely to commute on public transportation, and the commute times are typically long. The socioeconomic composition is stratified in terms of education and income. At the same time, high rates of unemployment, public assistance, and poverty are also common. Minority residents predominate, especially those who are Spanish-speaking. The concentration of children and young adults is high in these communities, as is the likelihood of female-headed households and large (six-person or greater) households.

Neighborhood Type 5: High SES, Foreign-Born, New Home Owners

The fifth neighborhood archetype has built environment characteristics indicative of a mixed urban-suburban area, with relatively low walkability, neither high nor low levels of density, and the third highest likelihood of being classified as mixed urbanicity. Residents are highly likely to be home owners and are also very likely to live in new housing. The communities are most clearly defined by the high value of housing stock. Housing turnover is low in the community, but the residents tend to be from out of state (i.e.stable migrants). People living in these communities are the most likely to have long commute times, and they also have a high likelihood of working at home. The socioeconomic status of these communities is the highest of all the neighborhood archetypes, particularly in terms of educational composition and income. This is the only neighborhood archetype in which more than half of the neighborhoods can be categorized as “high income.” In parallel, these neighborhoods are thus also the least likely to have high rates of poverty, public assistance, unemployment, or males out of the labor force. The age composition is concentrated in midlife (age 35–64), with the lowest likelihood of female-headed households. Demographically, the neighborhood archetype is also distinguished by a high prevalence of foreign-born residents.

Neighborhood Type 6: Middle-class Suburban and Exurban Families

The sixth neighborhood archetype is distinguished by built environment characteristics indicative of suburban and ex-urban neighborhoods. In addition, these neighborhoods have the highest likelihood of home ownership and the highest likelihood of new housing. The homes in these neighborhoods are the most likely of all of the classes to be valued in the middle of the housing stock distribution. Socioeconomically, the neighborhoods are middle-class, with mid-to high income and education. They are the most likely to be predominantly white, and the age composition suggests that they are comprised of families—with the highest concentration of children (population age 0–17) and a high concentration of adults in midlife (population age 35–64).

Sensitivity analyses

The findings on neighborhood characteristics presented above come from the findings of the final multigroup LCA. This final multigroup LCA was fit after identifying the best measurement model for 1990 and 2000, and determining that this model was qualitatively and quantitative consistent in terms of the number of neighborhood archetypes and their characteristics[1]. Specifics of the refinement of the measurement models that were conducted in the process of identifying the final, best fitting multigroup LCA model for the neighborhood characterization are described in Appendix 2. Sensitivity analyses show that the models are robust to alternative specifications of the characteristics included and their parameterization, and they show that the final model produces predicted

[1] The findings of the LCA measurement models for 1990 and 2000 are available upon request. Differences in the strength of specific characteristics are observed for neighborhood type 5 (with urbanicity and commuting more predominant in 1990 and home ownership and new housing in 2000) and in type 1 (which was predominantly young in 1990 but less so in 2000). Also, in 1990 there is a class of widows (single-household elderly), while in 2000 there is a class of foreign-born urban commuters.

neighborhood assignments that strongly differentiate between neighborhood archetypes (See Appendix 3). The final set of indicators and their parameterization is detailed in the third column of Table 1.

How does the prevalence of neighborhood archetypes in the U.S. change between 1990 and 2000?

The LCA approach allows us to assess how the prevalence of the six neighborhood archetypes changes between 1990 and 2000. For this assessment we hold constant the measurement of neighborhood archetypes (in terms of number and characteristics described above), and test whether the prevalence of neighborhood archetypes changes in a statistically significant way over the ten years.

Table 4 shows that there is a moderate 10-year change in the prevalence of each neighborhood archetype, and that this change in prevalence is (with two exceptions – type 3 and type 6) statistically significant for all neighborhood archetypes. Moreover the changes in the prevalence entail a re-ordering of the most frequently observed neighborhood archetypes. Specifically, type 5 (High SES, foreign born, new home owners) shifts from being the 4th most prevalent neighborhood archetype in 1990 to joining type 1 (Mobile, single-household, urbanites) as being the most prevalent neighborhood archetype in 2000. Additionally, type 2 (Low SES, rural) drops from being the 2nd *most* prevalent archetype in 1990 to the 2nd *least* prevalent archetype in 2000. Furthermore, increases in the prevalence of type 4 (Low SES, urban, minority commuters) are also observed.

How do individual neighborhoods changes shape neighborhood life cycles in the population?

Using the findings from the LCA model, we are able to examine how individual neighborhoods change between 1990 and 2000. This allows us to explore whether changes in individual neighborhood characteristics (e.g. an increase in minority residents in a specific neighborhood) have taken place so that the neighborhood is now better characterized by a different neighborhood archetype. This analysis thus provides information that can be used to understand the life cycle of neighborhoods as they move through different neighborhood archetypes.

The life cycle of stability and change observed for U.S. neighborhoods over the period 1990 to 2000 is summarized in Table 5. Neighborhood archetypes are predicted for all census tracts in the U.S. observed in 1990 and 2000 on the basis of posterior probabilities from the multigroup LCA detailed in Appendix 3 and Table 4. The distribution of the predicted neighborhood archetype in 2000 is displayed by the predicted neighborhood archetype in 1990. Thus, table 4 shows the proportion of neighborhoods that did not change archetypes between 1990 to 2000, or remained stable (i.e., along diagonal), and the proportion of neighborhoods that did experience a change in neighborhood characteristics that involve it being recategorized as a new neighborhood archetype (i.e., the off-diagonals).

Overall, stability is common with 75.4% of the individual neighborhoods classified as the same type between 1990 and 2000. The Kappa statistic for correspondence in neighborhood type characterization for individual census tracts between 1990 and 2000 shows moderate to low levels of change between archetypes over the period (Kappa=0.703).

The neighborhoods that do experience change (24.6% of the total), show a pattern of shifts from one neighborhood archetype to another (as summarized in Table 5) that extends the understanding of the overall pattern of change in the prevalence of neighborhood archetypes (in Table 4). Notably, even in the context of little, or no, change in the prevalence of neighborhood types in the total population, we observe that individual neighborhoods change from one neighborhood archetype to. Recall from Table 4 that neighborhoods of type 3 (Poor, urban, minority) and type 6 (Middle-class suburban/exurban families) had the least change in prevalence. Table 5 shows that the total prevalence of type 6 in the U.S. population was stable between 1990 and 2000, while Table 6 shows that that classification of individual neighborhoods as type 6 was unstable. In fact, individual neighborhoods classified as type 6 were the *most* likely of all of the neighborhoods to change archetypes between 1990 and 2000 (63.1 % correspondence). Thus, the instability for type 6 shown in Table 6, taken in the context of the stability for type 6 shown in Table 5, demonstrates that “flows” of neighborhoods in and out of type 6 occurred between 1990 and 2000, but that the flows of neighborhoods into type 6 were equal in number to the flows out of type 6.

Changes in and out of the neighborhood archetypes provide information about how the neighborhoods change over time—or in other words, information about neighborhood life cycle patterns. The top three largest changes between archetypes (comprising nearly 1/3 of all observed) are from type 2 (Low SES, rural) to type 6, from type 1 (Mobile single-household, urbanites) to type 4, and from type 4 (Low SES, urban, minority commuters) to type 3 (Poor, urban minority). Furthermore, for types 2 and 6, the flows 'out' are more than twice the flows 'in'. This is also the case for flows in and out of type 1 and type 4.

Thus, the findings from Table 5 show that the most change is observed for individual neighborhoods categorized as type 6 in 1990, followed by those in type 1, type 4 and type 2. These changes involve flows between archetypes that indicate neighborhood 'life cycle change' from "Low SES, rural" to "Middle-class suburban/exurban families" (e.g. for types 2 and 6), and from "Mobile single-household, urbanites" to "Low SES, minority, urban, commuters" (for types 1 and 4).

DISCUSSION

The primary objective of this study was to study neighborhood characterization and neighborhood change using a neighborhood archetype approach and latent class analysis (LCA) methodology. We studied the structure and change in neighborhood archetypes in the U.S. between 1990 and 2000 as a demonstration of our approach, observing the following principal findings. There are six different archetypes of neighborhoods in the U.S. characterized similarly in 1990 and 2000 by distinct sets of characteristics in the social and built environment, the migration and commuting patterns, and demographics and household patterns:

- Type 1: Mobile single-household, urbanites; Type
- Type 2: Low SES, rural
- Type 3: Poor, urban, minority
- Type 4: Low SES, urban, minority commuters
- Type 5: High SES, foreign born, new home owners
- Type 6: Middle-class suburban/exurban families

Between 1990 and 2000 the distribution of these neighborhood archetypes in the U.S. population changed in a small but statistically significant way, with notable increases in type 5 and type 4 and decreases in type 2 and type 1. Accompanying this distributional change was a moderate change by individual neighborhoods between the archetypes, as 24.6% experienced changes in population composition, migration and commuting or the built environment that involved the neighborhood being recategorized as a different archetype. The predominant patterns of change for individual neighborhoods (e.g. shifts from type 2 to type 6 and shifts from type 1 to type 4) indicate neighborhood life cycle dynamics consistent with urbanization (and ex-urbanization), both gentrification and neighborhood decline, Hispanic immigration and urban concentration, as well as the emergence of community structures configured by commuting patterns.

The innovation of this study is to employ the LCA methodology to advance substantive and methodological research on neighborhood characterization and neighborhood change in the U.S., and to do so in a way most relevant to population health research.

For researchers, the neighborhood archetypes approach and LCA method offers an efficient and statistically robust means of summarizing the combination of interacting conditions that constitute neighborhood risks and resources. We find that neighborhood archetypes are distinct constellations of characteristics across the domains of the built environment, migration and commuting, socioeconomic composition, and demographics and household composition. This is reflected by the way in which different groups of the archetypes are distinguished from one another on the basis of urbanization, race/ethnicity, class and family life cycle. It is also reflected in the interacting synthesis of community conditions, structure and population flows that emerge through our comparison between the neighborhood typology and the recent discourse on the 'post industrial city'.

Our findings for the U.S. on neighborhood characterization and change are consistent with research on changes in neighborhood composition by race, ethnicity, and social class that reflect residential segregation, gentrification and urban decline (e.g., Jargowsky 1997; Iceland et al. 2005). For example, type 3 and type 4 (but notably type 3) are consistent with the conditions of concentrated disadvantage that researchers studying health and social well-being in the inner-city have linked with deindustrialization, job loss, community deterioration, and both class and race-based 'flight' (e.g., see Wilson 1987). Both (but notably type 4) are also consistent with the focus of a large body of recent research on migration, assimilation, and acculturation (i.e. see, Logan et al. 2002; Fong & Shibuya 2005). The neighborhood archetypes also reflect distinctions between neighborhoods that have been discussed in research on the post-industrial globalized city (i.e. see Logan & Molotch 1987; Marcuse 1997; Nijman 2000). For example type 6 is consistent with the "edge-city communities of the middle class" described by Marcuse (1997, 2000) as postindustrial extensions of neighborhoods that support nuclear family life styles. Indeed, the growth in type 6 that we observed is consistent with the economic and technological transformations that have reconfigured work, home and family life (Bird and Rieker 2008). These are innovations which have allowed people to live in places where they can optimize the distances between where they want to work, raise families, and spend their leisure time. Furthermore, the patterns are consistent with theory on how neighborhoods themselves experience life course changes in their characteristics and composition over time (e.g. reviewed by Robert et al forthcoming). The consistency between our findings and those in prior quantitative and ethnographic work in community and urban studies provide us with confidence that our archetypes have theoretical validity.

The neighborhood archetypes approach and the resulting neighborhood measures that derive from this approach offers substantive insights to neighborhood characterization that are unlikely to be reflected using 'variable approaches' like factor analysis and structural equation modeling. While the continuous variables produced by these alternative methods (e.g. NSES) can be categorized into a nominal scale, factor analysis and structural equation models provide no statistical guidance on where and how to categorize the underlying continuum to produce archetypes. Furthermore, the most commonly employed alternative for considering archetypes -- cluster analysis—also has disadvantages relative to LCA reviewed in detail by Hagenaars and colleagues (1989). Thus, previous methodological research has found that LCA offers a more statistically rigorous method for classifying archetypes (in this case neighborhoods) on the basis of observed characteristics (Rapkin et al. 1993; Chow 1998).

Using LCA to identify neighborhood archetypes has allowed us to identify distinctions between neighborhoods that are not captured on continuous scales such as those arraying neighborhoods on the basis of socioeconomic advantage and disadvantage. For example, like Marcuse (1997) we find neighborhoods that are advantaged (type 1, type 5 and type 6), and disadvantaged (type 2, type 3 and type 4). But within these advantaged and disadvantaged communities there are factors like urbanicity, immigration, suburbanization, labor market involvement, transportation patterns, and the organization of the home-workplace balance that further differentiate between the neighborhood archetypes.

As throughout we emphasize the contribution that the study of neighborhood archetypes and health either statically or dynamically may offer for understanding health and health disparities. With respect to the dynamics of neighborhood change, we suspect that the neighborhood archetypes we have developed here may be particularly illuminating when used to studying questions regarding aging in place and social disparities in health among older adults.

Ecological approaches to neighborhood change such as the invasion-succession models and the neighborhood life cycle models have been used to help illuminate the environmental conditions influencing social deviance and poor human development (e.g., Bursik & Webb 1982). However to our knowledge, no study to date has examined in detail the role of accumulating exposure (e.g. the accumulated number of years) people are exposed to different neighborhood contexts on their health. This is a promising area in life course research (Sampson 2002; Robert et al. forthcoming) that could be facilitated with the use of neighborhood archetypes, such as those developed here.

CONCLUSION

This study is designed to both demonstrate the flexibility of the approach to a large number of indicators, and also to produce a construct which is a substantive contribution to the neighborhood and health literature. We underscore that the modeling conducted here is an illustration of the benefits and opportunities for further research made possible when taking a neighborhood archetypes approach. It is beyond the scope of this paper to develop the "best" neighborhood model. However we contend that we have demonstrated an important methodological advance in neighborhood research that opens new opportunities for further research, particularly for expanded time-periods in the U.S. and in other countries. Further research should extend these findings to look at different population health outcomes at an individual and aggregated level in relationship to the neighborhood archetypes. Our findings can thus provide the basis for future studies addressing persisting 'basic science' questions about neighborhoods and health—such as the multiplicity of neighborhood dimensions, and the changing dynamics of neighborhoods over time. In addition, they provide a basis for future applied population health objectives in benchmarking, surveillance and targeting of neighborhood-level interventions by public health practitioners and policymakers measures.

References

Boardman, J.D., Saint Onge, J.M., Rogers, R.G., & Denney, J.T. (2005) "Race differentials in obesity: The impact of place", *Journal of Health and Social Behavior* 46 (3):229-243.

Bollen, K.A. (2002) "Latent variables in psychology and the social sciences", *Annual Review of Psychology* 53 605-634.

Browning, C.R. & Cagney, K.A. (2002) "Neighborhood structural disadvantage, collective efficacy, and self-rated physical health in an urban setting", *Journal of Health and Social Behavior* 43 (4):383-399.

Bursik, R.J. & Webb, J. (1982) "Community change and patterns of delinquency", *American Journal of Sociology* 88 (1):24-42.

Chow, J. (1998) "Differentiating urban neighborhoods: A multivariate structural model analysis", *Social Work Research* 22 (3):131-142.

Cohen, D.A., Inagami, S., & Finch, B. (2008) "The built environment and collective efficacy", *Health & Place* 14 (2):198-208.

Collins, C.A. & Williams, D.R. (1999) "Segregation and mortality: The deadly effects of racism?", *Sociological Forum* 14 (3):495-523.

Corburn, J. (2004) "Confronting the challenges in reconnecting urban planning and public health", *American Journal of Public Health* 94 (4):541-546.

Cummins, S., Curtis, S., Diez-Roux, A.V., & Macintyre, S. (2007) "Understanding and representing 'place' in health research: A relational approach", *Social Science & Medicine* 65 (9):1825-1838.

Cutter, S.L., Mitchell, J.T. & Scott, M. S. (2000) "Revealing the vulnerability of people and places: A case study of Georgetown County, South Carolina", *Annals of the Association of American Geographers* 90 (4):713-737.

Daniels, M.J., Dominici, F., Samet, J.M., & Zeger, S.L. (2000) "Estimating particulate matter-mortality dose-response curves and threshold levels: An analysis of daily time-series for the 20 largest US cities", *American Journal of Epidemiology* 152 (5):397-406.

Diez Roux, A. V. (2001) "Investigating neighborhood and area effects on health", *American Journal of Public Health* 91 (11):1783-1789.

Fong, E. & Shibuya, K. (2005) "Multiethnic cities in North America", *Annual Review of Sociology* 31 285-304.

Frumkin, H. (2003) "Healthy places: Exploring the evidence", *American Journal of Public Health* 93 (9):1451-1456.

Galea, S. & Vlahov, D. (2005) "Urban health: Evidence, challenges, and directions", *Annual Review of Public Health* 26 341-365.

Haan, M., Kaplan, G.A., & Camacho, T. (1987) "Poverty and health - Prospective evidence from the Alameda County Study", *American Journal of Epidemiology* 125 (6):989-998.

Hagenaars, J.A. & Halman, L. C. (1989) "Searching for ideal types: The potentialities of latent class analysis", *European Sociological Review* 5 (1):81-96.

Hamer, M. & Chida, Y. (2008) "Active commuting and cardiovascular risk: A meta-analytic review", *Preventive Medicine* 46 (1):9-13.

Iceland, J., Sharpe, C., & Steinmetz, E. (2005) "Class differences in African American residential patterns in US metropolitan areas: 1990-2000", *Social Science Research* 34 (1):252-266.

Janson, C.G. (1980) "Factorial social ecology - An attempt at summary and evaluation", *Annual Review of Sociology* 6 433-456.

Jargowsky, P.A. 1997, *Poverty and Place; Ghettos, Barrios, and the American City* Russell Sage Foundation, New York.

Jelleyman, T. & Spencer, N. (2008) "Residential mobility in childhood and health outcomes: a systematic review", *Journal of Epidemiology and Community Health* 62 (7):584-592.

Leyden, K.M. (2003) "Social capital and the built environment: The importance of walkable neighborhoods", *American Journal of Public Health* 93 (9):1546-1551.

Logan, J.R., Alba, R.D., & Zhang, W.Q. (2002) "Immigrant enclaves and ethnic communities in New York and Los Angeles", *American Sociological Review* 67 (2):299-322.

Logan, J.R. & Molotch, H.L. 1987, *Urban Fortunes: the Political Economy of Place* University of California Press, Berkeley, CA.

Mplus Version 4.2. 2006. Los Angeles, CA.

Marcuse, P. (1997) "The enclave, the citadel, and the ghetto - What has changed in the post-Fordist US city", *Urban Affairs Review* 33 (2):228-264.

Marcuse, P. & van Kempen, R. 2000, *Globalizing cities: A new spatial order,* Blackwell, Oxford, UK.

Nijman, J. (2000) "The paradigmatic city", *Annals of the Association of American Geographers* 90 (1):135-145.

Pandey, S., Coulton, C. 1994. "Unraveling neighborhood change using two-wave panel analysis: A case study of Cleveland in the 1980s", *Social Work Research* 18(2):83-96.

Ponce, N.A., Hoggatt, K.J., Wilhelm, M., & Ritz, B. (2005) "Preterm birth: The interaction of traffic-related air pollution with economic hardship in Los Angeles neighborhoods", *American Journal of Epidemiology* 162 (2):140-148.

Poortinga, W. (2006) "Perceptions of the environment, physical activity, and obesity", *Social Science & Medicine* 63 (11):2835-2846.

Rapkin, B.D. & Luke, D.A. (1993) "Cluster-analysis in community research - Epistemology and practice", *American Journal of Community Psychology* 21 (2):247-277.

Robert, S.A., Cagney, K.A., & Weden, M.M. (forthcoming) "A life course approach to the study of neighborhoods and health," in *Handbook of Medical Sociology*, Sixth Edition, Chloe E. Bird, Peter Conrad, & Allen M. Fremont, eds., Prentice Hall, New Jersey.

Sampson, R. J., Morenoff, J. D., & Gannon-Rowley, T. (2002) "Assessing "neighborhood effects": Social processes and new directions in research", *Annual Review of Sociology* 28 443-478.

Schwirian, K. P. (1983) "Models of neighborhood change", *Annual Review of Sociology* 9 83-102.

Singer, B., Ryff, C. D., Carr, D., & Magee, W. J. (1998) "Linking life histories and mental health: A person-centered strategy", *Sociological Methodology, Vol.28 1998* 28 1-51.

Srinivasan, S., O'Fallon, L. R., & Dearry, A. (2003) "Creating healthy communities, healthy homes, healthy people: Initiating a research agenda on the built environment and public health", *American Journal of Public Health* 93 (9):1446-1450.

Taaffe, E. and H. Gauthier (1973). *Geography of Transportation*. Englewood Cliffs, N.J., Prentice-Hall, Inc.

Vlahov, D. & Galea, S. (2002) "Urbanization, urbanicity, and health", *Journal of Urban Health-Bulletin of the New York Academy of Medicine* 79 (4):S1-S12.

Weden, M. A., Carpiano, R. A., & Robert, S. A. (2008) "Subjective and objective neighborhood characteristics and adult health", *Social Science & Medicine* 66 (6):1256-1270.

Wilson, J. Q. & Kelling, G. (1982) "The police and neighborhood safety: Broken windows", *The Atlantic Monthly* 127 29-38.

Wilson, W. J. 1987, *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy* University of Chicago Press, Chicago.

Wood, L. & Giles-Corti, B. (2008) "Is there a place for social capital in the psychology of health and place?", *Journal of Environmental Psychology* 28 (2):154-163.

Wyly, E. K. (1999) "Continuity and change in the restless urban landscape", *Economic Geography* 75 (4):309-338.

Table 1. Sample Characteristics, 1990 & 2000

	1990 Median, (freq.)	2000 Median, (freq.)	Characterization of variables used in final LCA model
Sample size	61258	63592	
20% Sample size	12252	13261	
<u>*Built Environment*</u>			
Density [population / km^2]	691.3	837.6	0 if density < median, else 1
% Rural	0.0	0.0	
Urbanicity (from % Rural)			
Urban (low % rural)	(58.79)	(62.07)	0 if % rural = 0%
Mixed (high and low % rural)	(22.00)	(25.03)	1 if % rural >0% & <100%
Rural (low % rural)	(19.22)	(12.90)	2 if % rural = 100%
Mean block size (ft^2)	992259.8	963410.7	
Walkability (gamma index)	0.442	0.436	
Walkability (nodes)	177	183	
Walkability (alpha index)	0.159	0.151	0 if alpha < median, else 1
Housing stock value ($ per house)	67037	103504	0 if $value < median, else 1
% Owner occupied dwellings	70.5	71.9	0 if % owner < median, else 1
% Vacant dwellings	6.8	6.1	0 if % vacant < median, else 1
Housing construction date (year)	1962	1968	0 if % vacant < median, else 1
	28.1	23.5	
Air quality (PM10 in ug/m^3)			
Air pollution (PM10 > 50 ug/m^3)	(3.41)	(1.11)	0 if PM10 ≤ 50, else 1
<u>*Migration & Commuting*</u>			
State residency (% born in state)	71.8	72.0	
Housing turnover (% different house five years earlier)	42.5	40.7	
% Different state five years earlier	6.5	6.0	
Residential mobility			
Low state residency, high turnover	(15.23)	(17.59)	0 if residency < median and turnover ≥ median
Low residency, low turnover	(33.65)	(32.10)	1 if residency < median and turnover < median
High residency, high turnover	(33.65)	(32.11)	0 if residency ≥ median and turnover ≥ median
High residency, low turnover	(17.47)	(18.20)	0 if residency ≥ median and turnover < median
% Short commute (<0.5 hrs)	72.5	67.2	0 if % short < median, else 1
% Medium commute (0.5-1.5 hrs)	26.0	30.0	0 if % medium < median, else 1
% Long commute (>1.5 hrs)	1.1	2.2	0 if % long < median, else 1
% Work at home	2.3	2.6	0 if % home < median, else 1
% Use vehicle to commute	90.7	91.5	
% Use public trans. to commute	1.1	1.1	0 if % public < median, else 1
% walk or bike to commute	2.8	2.0	0 if % walk/bike < median, else 1

(Continued next page)

Table 1 (Cont'd). Sample Characteristics, 1990 & 2000

	1990 Median, (or freq.)	2000 Median, (or freq.)	Characterization of variables used in final LCA model
Socioeconomic Composition			
% Less than high school education	24.4	18.0	
% High school diploma	30.8	29.6	
% Some college	17.7	20.5	
% BA or graduate studies	14.1	17.8	
Educational composition			
Mid-low	(49.33)	(49.31)	% under high school > median (and/or) % high school > median
Mid-high	(41.25)	(41.00)	% high school > median (and/or) % some college > median (and/or) % BA or more > median
Stratified	(9.42)	(9.69)	All other patterns of education distribution
% Not in labor force (females)	43.9	42.7	
% Not in labor force (males)	25.1	28.4	0 if % not in LF < median, else 1
% Unemployed	5.6	4.9	0 if % unemployed < median, else 1
% Public assistance	5.7	2.5	0 if % public assist. < median, else 1
% Less than poverty	7.7	7.2	0 if % poverty < median, else 1
% Income <$30,000	26.8	20.5	
% Income $30,000-$59,000	33.0	32.4	
% Income $60,000-$74,000	5.5	9.9	
% Income ≥$75,000	4.8	16.1	
Income composition			
Mid-low	(41.54)	(37.80)	% income less than 30K > median (and/or) % income $30-59K > median
Mid-high	(36.79)	(32.11)	% income $30-59K > median (and/or) % income $60-74K >median
High	(8.53)	(27.85)	% income $60-74K >median (and/or) % income $75K+ > median
Stratified	(13.14)	(2.24)	all other patterns of income distribution
Demographics and Household Composition			
% Black	2.1	2.7	0 if % black < median, else 1
% White	88.8	80.8	0 if % white < median, else 1
% Hispanic	9.0	12.5	
% American Indian/ Alaskan Native	0.2	0.2	
% Foreign-born	3.0	4.8	0 if % foreign-born < median, else 1
% Spanish-speaking households	2.8	4.6	0 if % speak Spanish < median, else 1
% Asian-speaking households	0.5	0.8	
% Linguistically isolated Spanish-speaking households	0.0	10.1	
% Linguistically isolated Asian-speaking households	0.0	11.2	

(Continued next page)

Table 1 (Cont'd). Sample Characteristics, 1990 & 2000

	1990 Median, (or freq.)	2000 Median, (or freq.)	Characterization of variables used in final LCA model
Demographics and Household Composition (Cont'd)			
% Children (population 0-17 yrs)	25.6	25.6	0 if % children < median, else 1
% Young adult (population 18-34 yrs)	26.0	22.0	0 if % young adult < median, else 1
% Midlife (population 35-64 yrs)	34.1	38.5	0 if % midlife < median, else 1
% Older adult (population 65-79 yrs.)	9.7	9.1	0 if % older adult < median, else 1
% Population 65+ yrs	12.5	12.2	
% Oldest adults (population 80+ yrs)	2.4	2.8	0 if % oldest old < median, else 1
% Singles (1-person household)	22.7	24.2	0 if % singles < median, else 1
% Large family (6-person household)	3.4	3.3	0 if % large family < median, else 1
% Female-headed household	7.0	6.1	0 if % female head < median, else 1

Table 2. Number of Neighborhood Archetypes in the U.S. in 1990 and 2000

Model	Bayesian Information Criteria	Entropy	Vuong-Lo-Mendell-Rubin Test		Number of Threshold Statistics
			Test= 2*Δ log-likelihood	(p-value)	
1990					
6-class	492793	0.921	6-class vs. 5-class= 8071	(0.045)	5
7-class	486166	0.916	7-class vs. 6-class= 5816	(0.110)	6
2000					
6-class	508382	0.921	6-class vs. 5-class=9630	(0.000)	10
7-class	502120	0.918	7-class vs. 6-class= 7243	(0.450)	10

Table 3. Characteristics of Neighborhood Archetypes in the U.S., by Substantive Domain

Domain	Type 1: Mobile single-household, urbanites	Type 2: Low SES, rural	Type 3: Poor, urban minority	Type 4: Low SES, urban minority commuters	Type 5: High SES foreign born, new home owners	Type 6: Middle-class suburban/ exurban families
Built environment	**Urban** Dense Walkable	**Vacancy** Home owner Poor housing **Rural**	**Poor housing** Vacancy **Walkable** **Urban** Dense	**Urban** **Dense** Walkable ***Air pollution***	**High quality housing** Home owner New housing	**Middle quality housing** **Home owner** **New housing** ***Mixed urbanicity***
Migration & Commuting	Public transport Short commute Walk/bike **Unstable migrant**	**Work at home** **Stable resident**	**Short commute** **Walk/bike** Public trans. ***Unstable resident***	**Public trans.** Long commute Walk/bike	**Long commute** Work at home ***Stable migrant***	
Socioeconomic Composition	Mid-low education	**Mid-low education** Poverty Male not in labor force Mid-low income Public assistance Unemployed	**Poverty** **Public assistance** **Unemployed** **Income mid-low** **Male not in labor force** Mid-low education	Unemployed Public assistance Poverty Mid-low education ***Stratified education*** ***Stratified income***	**Mid-high education** ***High income***	**Mid-high income**
Demographics & Household Composition	**Single-household** **Oldest adults** Foreign-born Older adults Young adults White	**Older adults** White Oldest adults Children	**Female-headed household** Single-household Black Large household Oldest adults Older adults Children Adolescents	**Spanish-speaking** **Young adult** Female-headed household **Large household** **Black** **Foreign born** Children	**Adults in midlife** Foreign-born White	**White** Adults in midlife **Children**

Notes: Findings summarized from multigroup LCA model detailed in Appendix 3. Characteristics with probability greater than 0.600 given class membership are included in the table and ordered in decreasing order of probability by domain. If the probability of a characteristic given latent class membership is the highest across all classes, it is denoted with boldtype. If the probability of a characteristic is the highest across all classes, but is not greater than 0.600, it is denoted in italic boldtype.

Table 4. Prevalence of Neighborhood Archetypes in the U.S.

Latent class type	Description*	1990 (%)	2000 (%)	Difference 2000-1990	Odds Ratio (2000/1990)
y=1	Mobile single-household, urbanites	21.0	19.0	-2.0	0.882***
y=2	Low SES, rural	18.3	14.5	-3.8	0.757***
y=3	Poor, urban, minority	16.8	16.9	0.1	1.007
y=4	Low SES, urban, minority commuters	14.4	16.4	2.0	1.166***
y=5	High SES, foreign born, new home owners	15.5	19.0	3.5	1.279***
y=6	Middle-class suburban/exurban families	13.9	14.2	0.3	1.025

Notes: Findings on the distribution of neighborhood archetypes (latent classes) and the test of the difference in the odds of class membership for 2000 relative to 1990 come from the structural component of the multigroup LCA model. The measurement model is detailed in Appendix 3. The goodness of fit statistics for the model are a BIC=1056817.983 and entropy =0.945. The tests of statistical significance are reported as: * p-value ≤ 0.050; ** p-value ≤ 0.010; *** p-value ≤ 0.001.

Table 5. Life Cycle of U.S. Neighborhood Archetypes Stability and Change from 1990 to 2000

	Stability	Change by Neighborhood Archetype in 2000						
	%	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6	Total
Neighborhood Archetype in 1990								
Type 1: Mobile single-household, urbanites (n=11,323)	73.78	73.78	0.98	8.45	10.47	4.46	1.85	100
Type 2: Low SES, rural (n=9,763)	77.42	1.89	77.42	5.76	0.48	1.29	13.15	100
Type 3: Poor, urban, minority (n=8,171)	81.68	6.28	3.48	81.68	8.05	0	0.51	100
Type 4: Low SES, minority, urban, commuters (n=7,211)	76.01	6.07	0.61	14.17	76.01	1.53	1.61	100
Type 5: High SES, foreign born, new home owners (n=7,712)	78.33	8.75	0.54	0.08	3.67	78.33	8.62	100
Type 6: Middle-class suburban/exurban families (n=6,460)	63.08	10.73	8.13	1.58	3.44	13.05	63.08	100
Total (n=50640)	75.40							

Notes: Neighborhood archetypes are predicted for all census tracts in the U.S. observed in 1990 and 2000 on the basis of posterior probabilities from the mulitgroup LCA detailed in Appendix 3 and Table 4. The distribution of the predicted neighborhood archetype in 2000 is displayed by the predicted neighborhood archetype in 1990.

Appendix 1. Methodological Detail for Latent Class Analysis

Statistical Model

The statistical formulation for the LCA model is equivalent to the formulation of a log-linear model for an incomplete frequency table (Hagenaars 1990; Vermunt 1997). To illustrate the structure of the LCA model, we describe a model in which there are only three indicator variables, instead of the thirty-four indicators in our final model. For example this might entail an indicator *a* for density, *b* for commuting via public transportation, and *c* for poverty. Using the notation of Vermunt (1997), we describe the LCA model structure as follows:

$$\pi_{yabc} = \pi_y \pi_{a|y} \pi_{b|y} \pi_{c|y} . \qquad \text{(Eq. 1)}$$

The probability π_{yabc} denotes the likelihood of the pattern of responses to the indicator variables *a, b, and c* for a specific census tract observation. The probability π_y is the likelihood of the respective class *y,* where *y* indicates the neighborhood archetype. This probability can be interpreted as the prevalence of the neighborhood archetypes. This part of the model is considered the structural component of the model.

The conditional probability $\pi_{a|y}$ describes the probability of indicator variable *a* given membership in neighborhood type *y.* The characterization of the neighborhoods is based on this relationship between the manifest indicator variables 'x' and the latent neighborhood type 'y' observed via the conditional probabilities $\pi_{x|y}$. For the model in Eq. 1, characterization of the attributes of each neighborhood types *y* is based on the set of these conditional probabilities of *a, b,* and *c*. These conditional probabilities are similar to factor loadings in that they demonstrate which characteristics are most descriptive of members of a given class. The LCA models used to characterize the neighborhoods in 1990 and 2000 separately employ the basic structure in Eq. 1, but instead of only three variables, they include all thirty-four indicator variables in Table 1.

After fitting the separate LCA models for each year, we fit a multigroup LCA model. This model is an extension of the model in Eq. 1 in that it holds the measurement model constant for each year, and expands the model to condition π_y on a dichotomous indicator for the year (2000 versus 1990). Thus the final multigroup LCA model we test allows us to assess whether the value of π_y for each 'y' changes between 1990 and 2000.

The LCA models are estimated using an expectation maximization algorithm for variance adjusted weighted least squares. This estimation approach allows us to address missing data on the observed indicator variables. In addition, all models are run using multiple seed values to ensure that the estimates of the parameter values have indeed converged at a maximum likelihood, as opposed to a "local maxima."

Overall model fit is assessed using the Lo-Mendell-Rubin likelihood ratio test, the Bayesian Information Criteria (BIC) and entropy measures. The Lo-Mendell-Rubin test entails comparing the fitted LCA model to one with one fewer latent classes (Lo et al. 2001). The BIC is a modified likelihood measure that penalizes for over-fitting in terms of the number of parameters in the model (Hagenaars, 1990). Entropy is a summary measure that uses the posterior probabilities to determine the degree to which the classification can be clearly distinguished using the data (see Ramaswamy et al., 1993). Decisions about refinement of the indicator variables (discussed below in Appendix 2) are based on statistical tests of significance of parameters in the measurement model, as well as qualitative evidence on substantive differentiation of the latent classes, and the changes in the overall model fit.

After the best fitting multigroup LCA model is identified, the posterior probabilities ($p(y|x)$) are used to predict the latent class membership for all observations in the dataset.

Appendix 2. Refinement of the Latent Class Analysis Models and Sensitivity Analysis

Refinement of Latent Class Analysis Models

In order to identify the best fitting measurement model, we used qualitative and quantitative evidence to substantiate decisions about the inclusion and parameterization of indicator variables. After changes in the measurement model were made, the structural model was reassessed to determine if changes in the number or parameterization of the indicator variables influenced the number of classes.

We first determined whether changes in the parameterization of indictors improved overall model fit and the differentiation between latent classes. Indicators for closely related concepts were combined into a single variable. In specific, the following combinations were produced: state residency and housing turnover were combined into a single indicator for residential mobility pattern; educational indicators (less than high school, high school, some college, BA or graduate) were combined into a single indicator for educational composition pattern; and income indicators (less than $30,000, $30,000 to $59,000, $60,000 to $74,000, and greater than or equal to $75,000) were combined into a single indicator for income composition pattern.

We then assessed whether any indicators could be trimmed. We removed indicators that did not help to improve the differentiation of neighborhood archetypes, as observed through comparison of the conditional probabilities, $\pi_{x|y}$. Due to the strong statistical significance of all indicators, qualitative considerations were primary (e.g. see standard errors in Appendix 3). Through trimming, we observed that, among the measures of walkability, the alpha index provided the best qualitative differentiation between neighborhood archetypes. Similarly, among the commuting indicators, measurement of the use of a vehicle did not help to further distinguish between classes when the time spent commuting and the use of public transportation and walking or biking were incorporated. For the labor force variables, indicators for both men and women were observed to be redundant; men's labor force participation was the strongest of these two indicators. For race, ethnicity and language variables, we determined that indicators for % black, % white, % foreign-born, and % Spanish-speaking households provided the best model differentiation. None of the linguistic isolation variables improved qualitative differentiation of neighborhood archetypes, and Hispanic ethnic origin was less robust than the indicator for Spanish-speaking households. Finally, with respect to the age composition, we found that a single indicator for 65+ was not sufficient to capture the demographic composition of older adults. Models with two indicators for older adults—one capturing age 65-74 and one for age 80 and older-- produced more nuanced findings.

Sensitivity Analysis

The final six-class measurement models were robust to a number of alternative specifications. First, we examined a 10 percent sample of the census in each year using the indicator variables in their original continuous-variable functional, rather than the dichotomous and categorical variables created for the LCA. Using k-median and k-mean cluster analysis, we found that the number of groups was equal to the number of classes in our LCA models.[2] Although we found the same number of groups using cluster analysis as we did with the LCA, we found that the cluster analysis was more heavily influenced by the functional form of the indicator variables than the LCA and we found that the distribution of census tracts across groups was more skewed with the cluster analysis than with the LCA[3]. In addition, by using cluster analysis we were unable to categorize neighborhoods with missing data, such that 3 percent of the neighborhoods (205 tracts from the 1990 20% sample) were not classified with the cluster analyses that were classified with the LCA analyses

[2] We found that the Calinski/Harabasz pseudo-F statistic was minimized for the six-group cluster model indicating that six groups produced the best model. Findings available from authors upon request.
[3] Findings available from authors upon request.

Next, we examined the sensitivity to neighborhood characterization by considering measurement models with alternative variable specifications. We considered models in which each of the individual variables comprising the categorical composites (i.e., the variables for the migration pattern, educational composition, and income composite) were entered as individual dichotomous indicators. Using these models, we found that a seven-class structure best fit the data. These seven-class structures were identical to the seven-class models in all characteristics; however, they had an additional very low probability class in which all of the indicator variables were observed. The entropy statistics were poor for these models, and the models were difficult to interpret. For this reason, we found that our models were most robust with composite indicators of the selected variables.

Subsequently, to assess the sensitivity of the LCA models to the data sample, we estimated the models in different 20 percent and 10 percent samples of the data. These analyses reflected the potential danger of conducting LCA without extensive sensitivity analysis. In one of these samples, we produced a qualitatively and quantitatively different model of the neighborhood archetypes from all of the other samples. If we had use this model alone and had not conducted the sensitivity analyses with multiple samples, our findings may have been substantially skewed. Instead, with the exception of this one errant model, we have found marked consistency across all the samples and, thus, have much stronger confidence in the reliability of our neighborhood typologies.

Finally, we evaluated how well the LCA characterized the census tracts into distinct neighborhood archetypes on the basis of the posterior probabilities from the LCA models. By examining the average predicted probability across census tracts for each of the neighborhood archetypes given, we can identify how well our models distinctly classify the observations into different neighborhood archetypes. We can also identify the likelihood of being classified into alternative neighborhood archetypes. Overall, the neighborhood archetypes are well distinguished from other neighborhood archetypes in 1990 and 2000.

This is reflected in the finding that they all have a high (above 0.900) average probability of membership for the designated class[4].

[4] Findings available from authors upon request.

Appendix 3. LCA Model for Neighborhood Archetypes and Changes in the Distribution of Archetypes from 1990 to 2000, Model Results Reported in Conditional Probability Scale

Characteristic '*x*'	Probability of characteristic 'x' given latent class 'y', $\pi_{x\|y}$ (standard error) *					
	y=1	*y=2*	*y=3*	*y=4*	*y=5*	*y=6*
Built Environment						
Density (x_1)	0.841 (0.013)	0.003 (0.001)	0.621 (0.013)	0.850 (0.013)	0.416 (0.015)	0.160 (0.011)
Urbancity **						
Urban ($x_{2=1}$)	0.921	0.017	0.788	0.921	0.596	0.263
Mixed ($x_{2=2}$)	0.073	0.283	0.195	0.067	0.272	0.473
Rural ($x_{2=3}$)	0.006	0.700	0.017	0.012	0.132	0.265
Walkability ($x_{4=1}$)	0.729 (0.010)	0.247 (0.008)	0.795 (0.008)	0.758 (0.012)	0.239 (0.012)	0.248 (0.010)
Home owner ($x_{5=1}$)	0.231 (0.017)	0.864 (0.009)	0.187 (0.010)	0.145 (0.010)	0.849 (0.014)	0.873 (0.010)
Vacancy ($x_{6=1}$)	0.369 (0.014)	0.879 (0.006)	0.838 (0.009)	0.412 (0.020)	0.212 (0.007)	0.343 (0.011)
Age of housing ($x_{7=1}$)	0.376 (0.011)	0.635 (0.009)	0.214 (0.012)	0.395 (0.014)	0.727 (0.011)	0.786 (0.010)
Air pollution ($x_{8=1}$)	0.016 (0.002)	0.010 (0.002)	0.009 (0.002)	0.074 (0.005)	0.021 (0.003)	0.006 (0.002)
Housing stock (poor, $x_{9=1}$)	0.102 (0.013)	0.778 (0.012)	0.905 (0.022)	0.172 (0.035)	0.000 (0.000)	0.010 (0.006)
Housing stock (middle, $x_{10=1}$)	0.445 (0.019)	0.197 (0.011)	0.094 (0.019)	0.365 (0.014)	0.000 (0.000)	0.986 (0.006)
Housing stock (high, $x_{11=1}$)	0.450 (0.030)	0.024 (0.003)	0.001 (0.003)	0.461 (0.027)	1.000 (0.000)	0.000 (0.000)

(Continued next page)

* For each respective characteristic 'x', the highest conditional probability of a latent class is shaded gray. For example if density is high, latent class 4 is most likely (probability is 0.850).

**Distribution across latent classes for nominal variables are obtained from predicted distribution of latent classes in the sample data averaged over the two years.

Appendix 3 (Cont'd). LCA Model for Neighborhood Archetypes, 1990-2000

Characteristic '*x*'	Probability of characteristic 'x' given latent class 'y', $\pi_{x\|y}$ (standard error) *					
	y=1	*y=2*	*y=3*	*y=4*	*y=5*	*y=6*
Migration & commuting						
Migration**						
Stable migrant ($x_{12=1}$)	0.115	0.149	0.106	0.176	0.299	0.153
Unstable migrant ($x_{12=2}$)	0.512	0.130	0.241	0.382	0.377	0.280
Stable resident ($x_{12=3}$)	0.234	0.624	0.251	0.242	0.251	0.411
Unstable resident ($x_{12=4}$)	0.139	0.098	0.402	0.200	0.073	0.147
Work at home ($x_{13=1}$)	0.451	0.747	0.394	0.263	0.659	0.548
	(0.012)	(0.010)	(0.014)	(0.015)	(0.012)	(0.014)
Short commute (x_{14}=1)	0.662	0.499	0.789	0.256	0.248	0.549
	(0.026)	(0.013)	(0.023)	(0.022)	(0.013)	(0.015)
Long commute (x_{15}=1))	0.354	0.484	0.340	0.746	0.756	0.465
	(0.025)	(0.013)	(0.024)	(0.021)	(0.013)	(0.015)
Public transport (x_{16}=1)	0.684	0.094	0.693	0.852	0.541	0.173
	(0.016)	(0.007)	(0.017)	(0.015)	(0.014)	(0.011)
Walk/bike (x_{17}=1)	0.602	0.560	0.747	0.651	0.202	0.208
	(0.015)	(0.012)	(0.012)	(0.015)	(0.012)	(0.012)
Socioeconomics						
Educational Comp.**						
Stratified (x_{17}=1)	0.118	0.048	0.102	0.196	0.035	0.061
Mid-low (x_{17}=2)	0.178	0.881	0.852	0.676	0.052	0.450
Mid-high (x_{17}=3)	0.704	0.071	0.046	0.128	0.914	0.489
Male not in LF (x_{18}=1)	0.494	0.782	0.856	0.446	0.153	0.190
	(0.019)	(0.010)	(0.010)	(0.025)	(0.013)	(0.013)
Unemployed (x_{19}=1)	0.290	0.640	0.909	0.848	0.108	0.263
	(0.017)	(0.011)	(0.012)	(0.026)	(0.008)	(0.013)
Public assist. (x_{20}=1)	0.239	0.729	0.967	0.839	0.083	0.202
	(0.019)	(0.011)	(0.006)	(0.025)	(0.007)	(0.013)
Poverty (x_{21}=1)	0.230	0.810	0.970	0.819	0.015	0.200
	(0.023)	(0.010)	(0.007)	(0.026)	(0.003)	(0.016)
Income Comp.**						
Stratified (x_{22}=1)	0.109	0.104	0.056	0.125	0.001	0.043
Mid-low (x_{22}=2)	0.217	0.763	0.908	0.455	0.002	0.088
Mid-high (x_{22}=3)	0.447	0.114	0.028	0.317	0.404	0.764
High (x_{22}=4)	0.226	0.020	0.084	0.103	0.593	0.106

(Continued next page)

* For each characteristic 'x', the highest conditional probability of a latent class is shaded.
**Distribution obtained from sample data averaged over the two years.

Appendix 3 (Cont'd). LCA Model for Neighborhood Archetypes, 1990-2000

Characteristic '*x*'	Probability of characteristic 'x' given latent class 'y', $\pi_{x\|y}$ (standard error) *					
	y=1	*y=2*	*y=3*	*y=4*	*y=5*	*y=6*
Demographics & Household Composition						
Black (x_{23}=1)	0.495 (0.020)	0.313 (0.012)	0.799 (0.015)	0.804 (0.011)	0.298 (0.012)	0.335 (0.014)
White (x_{24}=1)	0.547 (0.027)	0.700 (0.012)	0.335 (0.020)	0.023 (0.005)	0.642 (0.013)	0.774 (0.015)
Spanish-speaking (x_{25}=1)	0.492 (0.025)	0.254 (0.010)	0.576 (0.018)	0.944 (0.008)	0.516 (0.013)	0.361 (0.014)
Foreign born (x_{26}=1)	0.721 (0.022)	0.093 (0.007)	0.369 (0.012)	0.874 (0.022)	0.775 (0.011)	0.203 (0.013)
Pop. 0-17 yrs. (x_{27}=1)	0.051 (0.009)	0.642 (0.012)	0.651 (0.026)	0.719 (0.035)	0.450 (0.020)	0.721 (0.016)
Pop. 18-34 yrs. (x_{28}=1)	0.646 (0.022)	0.127 (0.009)	0.648 (0.015)	0.923 (0.008)	0.330 (0.014)	0.409 (0.016)
Pop. 35-64 yrs. (x_{29}=1)	0.473 (0.023)	0.589 (0.012)	0.203 (0.010)	0.163 (0.017)	0.908 (0.008)	0.753 (0.014)
Pop. 65-80 yrs. (x_{30}=1)	0.706 (0.022)	0.744 (0.012)	0.605 (0.028)	0.169 (0.020)	0.328 (0.020)	0.244 (0.017)
Pop. ≥80 yrs. (x_{31}=1)	0.731 (0.020)	0.696 (0.013)	0.660 (0.028)	0.247 (0.021)	0.265 (0.019)	0.207 (0.015)
Female-head (x_{32}=1)	0.474 (0.024)	0.341 (0.013)	0.966 (0.006)	0.884 (0.014)	0.115 (0.009)	0.266 (0.014)
One person household (x_{33}=1)	0.912 (0.012)	0.406 (0.013)	0.855 (0.020)	0.439 (0.024)	0.121 (0.024)	0.097 (0.012)
Six person household (x_{34}=1)	0.123 (0.015)	0.513 (0.012)	0.679 (0.028)	0.880 (0.028)	0.537 (0.017)	0.482 (0.016)

* ☐ For each characteristic 'x', the highest conditional probability of a latent class is shaded.
Note: goodness of fit statistics: BIC=1056817.983; entropy=0.945)