Published in final edited form as:
Health Place. 2011 January ; 17(1): 289–299. doi:10.1016/j.healthplace.2010.11.002.

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], and **Nicole Lurie, MD, MSPH**[3]

[1] RAND Corporation

[2] University of California, Los Angeles

[3] U.S. Department of Health and Human Services

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 health-related risks and resources. Research on neighborhood effects has reconnected public health with its earlier population foundations—showing that social ecology and built environments are important "upstream" determinants of health. 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 (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, measured primarily with Census data using factor analysis. The theory and methods also informed the most commonly employed measures of context 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 ecological context 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. 2010). We develop a new, complementary theoretical and methodological approach to study neighborhoods that employs archetypes to characterize neighborhoods 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, Built Environment, 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 (e.g., Sampson et al. 2002; Cummins et al. 2007). Latent construct approaches are advocated for the measurement and characterization of social, cultural, political and geographic cleavages that entail correlated and overlapping constructs that are typically not well captured by any one individual indicator (e.g., Weden et al. 2008; see also Janson 1980; Schwirian 1983).

Early and most commonly used measures of neighborhood conditions rate neighborhoods on a single continuous scale (e.g., using factor-based scales commonly developed from sociodemographic indicators like local rates of poverty, public assistance, female headed households, unemployment, and African American residents; for review see Browning & Cagney 2002). The most common of these scales extends to a neighborhood-level the integrated assessment of class, status, and prestige through neighborhood socioeconomic status (NSES). Recent studies have focused on the multi-dimensionality of place-based social stratification. For example, Weden and colleagues (2008) develop a two-dimensional latent model for neighborhood-based social stratification that demonstrates the independent relevance of neighborhood disadvantage and neighborhood affluence to individual health. This is consistent with other work highlighting the relationship between structural aspects of neighborhoods (e.g. boarded-buildings, vacancy rates, and residential turnover; Wilson & Kelling 1982) and social disorganization that emerges from 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).

From a theoretical perspective, research in urban planning, urban studies and social ecology provide a foundation for 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 showed that ‘neighborhoods matter’ and have independent effects beyond individual socioeconomic characteristics (for review see Robert 1999). 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. Theory on the multidimensional experience of local environments has been well developed by scholars of gender and race who employ the theoretical paradigm of "intersectionality" to describe the contingent and interacting dimensions of social stratification (e.g., Choo and Ferree 2010). And though direct attention to "intersectionality" has been addressed in only one previous known study on neighborhoods and health (Kershow and Forer 2010), the potential importance of the concept is illustrated in previous literature. 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 et al. 2000; Ponce et al. 2005).

To date, most work has employed a 'variable perspective' to consider the multidimensionality of place based social stratification. For example, Boardman and colleagues (2005) model the potentially contingent role of neighborhood poverty based on racial segregation by exploring the role of an interaction between the two variables. The problem of data intensiveness required when taking a variable approach to multidimensionality becomes evident when additional axes are considered (e.g., a simple model with single dichotomous indicators for each of the four axes of the social ecology model would require 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– an analytical problem analogous to that encountered by life course researchers seeking to assess interactions across multiple domains of individual experience (see Singer et al.1998).

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 by characterizing archetypes, and the empirical method of latent class analysis (LCA) designed for this purpose (e.g., Hagenaars & Halman 1989) has been used extensively in social, behavioral, and health research (Bollen 2002)[1]. 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. 2010).

[1]Conceptually similar, but alternative latent measurement approaches for modeling archetypes include latent profile analysis and grade of membership (Gibson 1959; Woodbury and Manton 1982).

Although latent measurement methods related to LCA have been developed to address issues of bias in population health research (e.g., see methods for addressing spatial clustering in the association between neighborhood deprivation and area-level health Congdon 1997, Congdon 1996a, Congdon 1996b), to our knowledge, LCA has not been applied to study both the characterization and change in neighborhoods across the range of social and built environment domains relevant to population health. This is unfortunate, since the approach offers distinct analytical advantages to alternative methods that have predominated the literature (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 offer opportunities to advance research on neighborhoods and health. First, LCA can measure how constellations of characteristics capture distinct neighborhood archetypes. LCA, like other latent measurement methods, offers the advantage of addressing uncertainty, bias, and potential attenuation due to systematic and stochastic error in the measurement of variables. Additionally, LCA allows a researcher to explore the constellations of characteristics that would otherwise need to be modeled using ‘interactions’ between neighborhood dimensions in a model of neighborhood and health. 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, and to do so causally external from the impact of the neighborhood characteristics on health. Secondly, like factor analytic methods, LCA allows one to assess the stability or change of neighborhood archetypes over time using a temporally stable measurement methodology. These neighborhood archetypes provide a mechanism for studying outstanding questions about neighborhood life cycles. In summary, at its minimum, LCA is a data reduction mechanism similar to cluster analysis, but which offers the advantages of more fully addressing the potential biases of measurement error. 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 database contains 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. 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 indicators 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. Appendix 1 details the specific categorization of each of the indicators. Refinement of the indicators was conducted in conjunction with refinement of the measurement models, and technical details are described elsewhere (Weden et al. 2010).

Built Environment—Urbanization is measured by *density* and *urbanicity* 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 street-intersections or "*nodes*," and two measures of *walkability* (see 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* is included as in indicator of environmental pollution, with a threshold for health compromising levels of particulate matter smaller than 10 micrometers (PM10) of 50 ug/m^3 (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 the *mode of transportation to work*.

Socioeconomic Composition—Socioeconomic composition captures the aspects of NSES, neighborhood affluence, and neighborhood disadvantage described earlier. The indicators entail *educational attainment*, *labor force characteristics*, and *economic characteristics*.

Demographic and Household Composition—The racial/ethnic composition and life-cycle stage of residents are highlighted in the social ecology model and are captured here by *race and ethnicity, native language* and *age,* as well as 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. This model is used to assess whether the distribution of neighborhood archetypes and their characterization changes over time. Using the model, we hold characterization of the neighborhood archetypes constant and test whether the distribution of neighborhoods across neighborhood archetypes changes between 1990 and 2000.

LCA models are fit to the observed data on built environment, migration and commuting, socioeconomic composition, and demographic and household composition (Appendix 1), thereby allowing us to characterize the unobserved latent variable for the neighborhood archetypes. Refinement of the structural component of the LCA models (e.g. the number of archetypes) and the measurement components of the LCA models (e.g. the characteristics of the 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 & McCutcheon 2002; Ramaswamy et al. 1993; Lo et al. 2001).

Mplus software version 4.2 (Mplus Version 4.2. 2006) allows us to conduct LCA accounting for missing observations on some variables. 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.

Statistical detail on the LCA modeling process, model parameterization and refinement of the structural and measurement models, as well as sensitivity analyses (including validation of the LCA model against findings from an alternative statistical method—cluster analysis) is described elsewhere (Weden et al. 2010).

FINDINGS

How many neighborhood archetypes are there in 1990 and 2000?

Six neighborhood archetypes best summarize the combinations of characteristics from the built environment, migration and commuting, socioeconomic composition, and demographics and household composition in the U.S. in both 1990 and 2000 (See Table 1). Specifically, a six-class LCA model produces the best goodness of fit statistics when models are fit to the data for each year, (see Weden et al. 2010 for statistical detail).

What are the defining characteristics of the neighborhood archetypes?

A summary of the combinations of characteristics for all of the neighborhood archetypes is provided in Table 1. This table qualitatively translates the statistical findings for the measurement of archetypes from the final multigroup LCA model that is fully detailed in Appendix 2[2]. Characteristics are reported by archetype (columns), organized by domain (rows), and reported in descending order of probability (within the cells). Finally, characteristics are denoted with bold archetype where the probability of the characteristic is highest across the archetypes. For example, there are three characteristics in the built environment domain that predominate for neighborhoods of type 1: urbanicity, density, and walkability. They are listed in descending order on the basis of the LCA findings for archetype 1 (see Appendix 2, Column 1): 0.921 probability of being categorized as urban; 0.841 probability of having density above the national median; and 0.729 probability that walkability is above the national median. Notice that the highest probability of being urban across all the archetypes is 0.921, and that both type 1 and type 4 are equally likely to have this highest likelihood of urbanicity, so urban is bolded in the list of build environment characteristics for both type 1 and type 4.

As follows, we describe the findings of the most prevalent neighborhood characteristics summarized in Table 1. We also note the least common characteristics that are detailed in Appendix 2 but not included in the Table 1 summary.

Neighborhood Type 1: Mobile Single-household, Urbanites—The first 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, type 1 are the least likely to have a higher than the national median prevalence of children (age 0–17 years) but the most likely to contain elderly adults (age 80 and older) and are also likely to contain young adults (age 18–34 years). Overall, these neighborhoods are most predominantly characterized by having the highest likelihood of single-person households.

[2]In Table 1, characteristics are listed for which the probability of the characteristic is greater than 0.6 in the class, as reported in Appendix 2.

Neighborhood Type 2: Low SES Rural—The second 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 is likely to have ‘mid-low’ income, and high levels of poverty, public assistance, unemployment, and males out of the labor force. The demographic composition is age-stratified. It has the highest likelihood of above median proportion of adults in the early retirement years (e.g., age 65–80) as well as a high proportion of young persons. It is predominantly white, with the lowest likelihood of black residents.

Neighborhood Type 3: Poor, Urban Minority—The third 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, 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). Socioeconomic disadvantage is high. On nearly all indicators of poor socioeconomics (i.e., poverty, public assistance, unemployment, income composition, and males out of the labor force), these neighborhoods are the most likely to exceed national norms. They also 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 archetype is among the two most urban and dense. It is also the most likely to have hazardous air quality, although the overall likelihood of hazardous PM10 levels across all six of the 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. Additionally, 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 archetype has built environment characteristics indicative of a mixed urban-suburban area, with relatively low walkability. It has neither high nor low levels of density, and it has 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 higher than national likelihoods of working at home. The socioeconomic status is the highest of all the archetypes, particularly in terms of educational composition and income. This is the only archetype in which more than half of the neighborhoods can be categorized as “high income.” In parallel, it is 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 archetype is also distinguished by a high prevalence of foreign-born residents.

Neighborhood Type 6: Middle-class Suburban and Exurban Families—The sixth archetype is distinguished by built environment characteristics indicative of suburban and ex-urban neighborhoods. It has the highest likelihood of home ownership and the highest likelihood of new housing. Homes in these neighborhoods are the most likely across the archetypes to be valued in the middle of the housing stock distribution. The archetype is middle-class, with mid-to high income and education. It is the most likely to be predominantly white, and the age composition suggests families—with the highest concentration of children (population age 0–17) and a high concentration of adults in midlife (population age 35–64).

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 2 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 archetypes. Moreover the changes in the prevalence entail a re-ordering of the most frequently observed archetypes. Specifically, type 5 (High SES, foreign born, new home owners) shifts from being the 2nd *least* prevalent archetype in 1990 to joining type 1 (Mobile, single-household, urbanites) as being the *most* prevalent archetype in 2000. Additionally, type 2 (Low SES, rural) drops from being the 2nd *most* prevalent archetype in 1990 to becoming the 2nd *least* prevalent archetype in 2000. Furthermore, increases in the prevalence of type 4 (Low SES, urban, minority commuters) are observed.

How do changes in characteristics of individual neighborhoods shape neighborhood life cycles in the U.S.?

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 shifts from one archetype to another. This analysis thus provides information that can be used to understand the life cycle of neighborhoods as their individual structure and composition changes over time.

The life cycle of stability and change observed for U.S. neighborhoods over the period 1990 to 2000 is summarized in Table 3. Neighborhood archetypes are predicted for all U.S. census tracts in each year (See Weden et al. 2010 for statistical detail), and their classification in 2000 is displayed by their classification in 1990. Thus, table 3 shows the proportion of neighborhoods that did not change archetypes and thus remained stable between 1990 to 2000 (along the diagonal), and the proportion of neighborhoods that did experience a change in neighborhood characteristics that involve it being recategorized as a new archetype (the off-diagonals).

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

The neighborhoods that do experience change (in Table 3), show a pattern of shifts from one archetype to another that extends the understanding of the overall pattern of change in the prevalence of archetypes in the U.S. described earlier (in Table 2). Notably, even in the context of little, or no, change in the prevalence of neighborhood archetypes across the U.S., we observe that individual neighborhoods change from one neighborhood archetype to another. Recall from Table 2 that neighborhoods of type 3 (Poor, urban, minority) and type 6 (Middle-class suburban/exurban families) had the least change in prevalence and thus greatest stability in its U.S. prevalence between 1990 and 2000. In contrast, Table 3 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 (1–63.1=46.9 % non-correspondence). Thus, the instability for type 6 shown in Table 3, taken in the context of the stability for type 6 shown in Table 2, 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 one-third of all observed) are from type 4 (Low SES, urban, minority commuters) to type 3 (Poor, urban minority), type 2 (Low SES, rural) to type 6, and type 1 (Mobile single-household, urbanites) to type 4. Furthermore, for types 2 and 6, as well as type 1 and type 4, the flows ‘out’ are more than the flows ‘in’.

In summary, the findings from Table 3 show that the neighborhoods that experiencing the most change via life cycle dynamics are those categorized in 1990 as type 6, followed by type 1, 4 and 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 of U.S. neighborhood archetypes between 1990 and 2000 as a demonstration of our approach, observing the following principal findings. There are six different archetypes that are characterized similarly across both years 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 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 archetypes changed in a small but statistically significant way, with notable increases in type 5 and type 4 neighborhoods and decreases in type 2 and type 1 neighborhoods. Accompanying this distributional change was a moderate change by individual neighborhoods from one archetype to another, illuminating neighborhood life-cycle dynamics over the ten year period. In total, 24.6% of the neighborhoods experienced such ‘life-cycle dynamics’ via changes in population composition, migration and commuting, or the built environment. 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 offer 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’ ((i.e. see Logan & Molotch 1987; Marcuse 1997; Nijman 2000).

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. 2010). 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 offer substantive insights to neighborhood characterization which are unlikely to be reflected using alternative latent 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 that is the model output. Nor do they allow researchers to follow individual’s migratory patterns across different types of neighborhood environments as is possible with categorical 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. Similarly, 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. 2010) 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 neighborhood archetype approach to a large number of social and environmental 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. For example, one of the previous critiques of neighborhood and health literature is the use of U.S. census tracts to proxy for actual or 'natural' neighborhood areas (Diez-Rouz 2001). Future research could employ the LCA approach developed here to compare the characterization of neighborhoods defined by alternative geographical boundaries which our data did not allow us to address. Similarly, although we expanded the range of neighborhood characteristics typically addressed in population health research, we made a theoretical and analytical decision to employ categorical and dichotomous indicators of these characteristics, describing local area experience in reference to national medians. Future research could explore additional or alternative indicators and parameterizations of these neighborhood indicators employing alternative latent methods for modeling archetypes like latent trait analysis or grade of membership (Gibson 1959; Woodbury and Manton 1982). Furthermore, we encourage future researchers to extend our findings by considering the relationship between this (or their own) neighborhood archetype model and actual population health outcomes.

These future applications and extensions of the neighborhood archetype approach developed here may identify nuances of population health relevant contextual conditions most relevant to specific local areas or health-related conditions. The findings of this study 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 space and time. Additionally, 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.

Acknowledgments

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, Erin Ruel, and our reviewers 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).

References

Boardman JD, Saint Onge JM, Rogers RG, Denney JT. Race differentials in obesity: The impact of place. Journal of Health and Social Behavior. 2005; 46(3):229–243. [PubMed: 16259146]

Bollen KA. Latent variables in psychology and the social sciences. Annual Review of Psychology. 2002; 53:605–634.

Browning CR, Cagney KA. Neighborhood structural disadvantage, collective efficacy, and self-rated physical health in an urban setting. Journal of Health and Social Behavior. 2002; 43(4):383–399. [PubMed: 12664672]

Bursik RJ, Webb J. Community change and patterns of delinquency. American Journal of Sociology. 1982; 88(1):24–42.

Choo HY, Feree MM. Practicing intersectionality in sociological research: A critical analysis of inclusions, interactions, and institutions in the study of inequalities. Sociological Theory. 2010; 28(2):129–149.

Chow J. Differentiating urban neighborhoods: A multivariate structural model analysis. Social Work Research. 1998; 22(3):131–142.

Cohen DA, Inagami S, Finch B. The built environment and collective efficacy. Health & Place. 2008; 14(2):198–208. [PubMed: 17644395]

Collins CA, Williams DR. Segregation and mortality: The deadly effects of racism? Sociological Forum. 1999; 14 (3):495–523.

Congdon P. Suicide and parasuicide in London: A small-area study. Urban Studies. 1996a; 33:137–158.

Congdon P. The Epidemiology of suicide in London. Journal of the Royal Statistical Society Series A. 1996b; 159(3):515–533.

Congdon P. Multilevel and clustering analysis of health outcomes in small areas. European Journal of Population. 1997; 13:305–338.

Corburn J. Confronting the challenges in reconnecting urban planning and public health. American Journal of Public Health. 2004; 94(4):541–546. [PubMed: 15053998]

Cummins S, Curtis S, Diez-Roux AV, Macintyre S. Understanding and representing 'place' in health research: A relational approach. Social Science & Medicine. 2007; 65(9):1825–1838. [PubMed: 17706331]

Cutter SL, Mitchell JT, Scott MS. Revealing the vulnerability of people and places: A case study of Georgetown County, South Carolina. Annals of the Association of American Geographers. 2000; 90(4):713–737.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Daniels MJ, Dominici F, Samet JM, Zeger SL. 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. 2000; 152(5):397–406. [PubMed: 10981451]

Diez Roux AV. Investigating neighborhood and area effects on health. American Journal of Public Health. 2001; 91(11):1783–1789. [PubMed: 11684601]

Downs, A. Neighborhoods and Urban Development. Brookings Institute; Washington, D.C: 1981.

Fong E, Shibuya K. Multiethnic cities in North America. Annual Review of Sociology. 2005; 31:285–304.

Frumkin H. Healthy places: Exploring the evidence. American Journal of Public Health. 2003; 93(9): 1451–1456. [PubMed: 12948962]

Galea S, Vlahov D. Urban health: Evidence, challenges, and directions. Annual Review of Public Health. 2005; 26:341–365.

Gibson WA. Three multivariate models: Factor analysis, latent structure analysis, and latent profile analysis. Psychometrika. 1959; 24:229–252.

Haan M, Kaplan GA, Camacho T. Poverty and health - Prospective evidence from the Alameda County Study. American Journal of Epidemiology. 1987; 125(6):989–998. [PubMed: 3578257]

Hagenaars JA, Halman LC. Searching for ideal types: The potentialities of latent class analysis. European Sociological Review. 1989; 5(1):81–96.

Hagenaars, JA.; McCutcheon, AL. Applied Latent Class Analysis Models. Cambridge University Press; Cambridge: 2002.

Hamer M, Chida Y. Active commuting and cardiovascular risk: A meta-analytic review. Preventive Medicine. 2008; 46(1):9–13. [PubMed: 17475317]

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

Janson CG. Factorial social ecology - An attempt at summary and evaluation. Annual Review of Sociology. 1980; 6:433–456.

Jargowsky, PA. Poverty and Place; Ghettos, Barrios, and the American City. Russell Sage Foundation; New York: 1997.

Jelleyman T, Spencer N. Residential mobility in childhood and health outcomes: a systematic review. Journal of Epidemiology and Community Health. 2008; 62(7):584–592. [PubMed: 18559440]

Kershaw P, Forer B. Selection of area-level variables from administrative data: An intersectional approach to the study of place and child development. Health & Place. 2010; 16(3):500–511. [PubMed: 20089438]

Leyden KM. Social capital and the built environment: The importance of walkable neighborhoods. American Journal of Public Health. 2003; 93(9):1546–1551. [PubMed: 12948978]

Lo Y, Mendell NR, Rubin DB. Testing the number of components in a normal mixture. Biometrika. 2001; 88(3):767–78.

Logan JR, Alba RD, Zhang WQ. Immigrant enclaves and ethnic communities in New York and Los Angeles. American Sociological Review. 2002; 67(2):299–322.

Logan, JR.; Molotch, HL. Urban Fortunes: the Political Economy of Place. University of California Press; Berkeley, CA: 1987.

Mplus Version 4.2. Los Angeles, CA: 2006.

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

Marcuse, P.; van Kempen, R. Globalizing cities: A New Spatial Order. Blackwell; Oxford, UK: 2000.

Nijman J. The paradigmatic city. Annals of the Association of American Geographers. 2000; 90(1): 135–145.

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

Ponce NA, Hoggatt KJ, Wilhelm M, Ritz B. Preterm birth: The interaction of traffic-related air pollution with economic hardship in Los Angeles neighborhoods. American Journal of Epidemiology. 2005; 162(2):140–148. [PubMed: 15972941]

Poortinga W. Perceptions of the environment, physical activity, and obesity. Social Science & Medicine. 2006; 63(11):2835–2846. [PubMed: 16952415]

Ramaswamy V, DeSarbo W, Reibstein D, Robinson W. An empirical pooling approach for estimating marketing elasticities with PIMS data. Marketing Science. 1993; 12(1):103–124.

Rapkin BD, Luke DA. Cluster-analysis in community research - Epistemology and practice. American Journal of Community Psychology. 1993; 21(2):247–277.

Robert SA. Socioeconomic position and health: the independent contribution of community socioeconomic context. Annual Review of Sociology. 1999; 21:489–516.

Robert, SA.; Cagney, KA.; Weden, MM. A life course approach to the study of neighborhoods and health. In: Bird, Chloe E.; Conrad, Peter; Fremont, Allen M., editors. Handbook of Medical Sociology. 6. Prentice Hall; New Jersey: 2010. p. 124-143.

Sampson, Robert; Groves, Byron W. Community structure and crime: Testing social disorganization theory. American Journal of Sociology. 1989; 94:774–802.

Sampson RJ, Morenoff JD, Gannon-Rowley T. Assessing “neighborhood effects”: Social processes and new directions in research. Annual Review of Sociology. 2002; 28:443–478.

Schwirian KP. Models of neighborhood change. Annual Review of Sociology. 1983; 9:83–102.

Singer B, Ryff CD, Carr D, Magee WJ. Linking life histories and mental health: A person-centered strategy. Sociological Methodology, Vol. 28 1998. 1998; 28:1–51.

Srinivasan S, O’Fallon LR, Dearry A. Creating healthy communities, healthy homes, healthy people: Initiating a research agenda on the built environment and public health. American Journal of Public Health. 2003; 93(9):1446–1450. [PubMed: 12948961]

Taaffe, E.; Gauthier, H. Geography of Transportation. Prentice-Hall, Inc; Englewood Cliffs, N.J: 1973.

Vlahov D, Galea S. Urbanization, urbanicity, and health. Journal of Urban Health-Bulletin of the New York Academy of Medicine. 2002; 79 (4):S1–S12. [PubMed: 12473694]

Weden MM, Bird CE, Escarce JJ, Lurie N. 2010Technical detail and appendices for a study of neighborhood archetypes for population health research. RAND Working Paper WR-805. http://www.rand.org/pubs/working_papers/WR805/

Weden MM, Carpiano RA, Robert SA. Subjective and objective neighborhood characteristics and adult health. Social Science & Medicine. 2008; 66(6):1256–1270. [PubMed: 18248865]

Wilson JQ, Kelling G. The police and neighborhood safety: Broken windows. The Atlantic Monthly. 1982; 127:29–38.

Wilson, WJ. The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy. University of Chicago Press; Chicago: 1987.

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

Woodbury MA, Manton KG. A new procedure for analysis of medical classification. Methods of Information in Medicine. 1982; 21:183–244.

Wyly EK. Continuity and change in the restless urban landscape. Economic Geography. 1999; 75(4): 309–338.

Table 1

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-high 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-person household** **Oldest adults** Foreign-born Older adults Young adults	**Older adults** White Oldest adults Children	**Female-headed household** Single-person household Black Large household Oldest adults Children Young adults Older adults	**Spanish- speaking** **Young adult** Female-headed household **Large household** **Foreign born** **Black** Children	**Adults in midlife** Foreign-born White	**White** Adults in midlife **Children**

Notes: Findings summarized from multigroup LCA model detailed in Appendix 2 that employs indicator variables detailed in Appendix 1. 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 2

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 2. 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.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 3

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 2. The distribution of the predicted neighborhood archetype in 2000 is displayed by the predicted neighborhood archetype in 1990.

Appendix 1

Neighborhood Indicator Variables and Sample Statistics, 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	
Built Environment			
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
Migration & Commuting			
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

	1990 Median, (freq.)	2000 Median, (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	
% 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

	1990 Median, (freq.)	2000 Median, (freq.)	Characterization of variables used in final LCA model
% 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

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Appendix 2

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)
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 (0.012)	0.747 (0.010)	0.394 (0.014)	0.263 (0.015)	0.659 (0.012)	0.548 (0.014)
Short commute ($x_{14=1}$)	0.662 (0.026)	0.499 (0.013)	0.789 (0.023)	0.256 (0.022)	0.248 (0.013)	0.549 (0.015)
Long commute ($x_{15=1}$))	0.354 (0.025)	0.484 (0.013)	0.340 (0.024)	0.746 (0.021)	0.756 (0.013)	0.465 (0.015)
Public transport ($x_{16=1}$)	0.684 (0.016)	0.094 (0.007)	0.693 (0.017)	0.852 (0.015)	0.541 (0.014)	0.173 (0.011)

NIH-PA Author Manuscript

NIH-PA Author Manuscript

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*
Walk/bike ($x_{17=1}$)	0.602 (0.015)	0.560 (0.012)	0.747 (0.012)	0.651 (0.015)	0.202 (0.012)	0.208 (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 (0.019)	0.782 (0.010)	0.856 (0.010)	0.446 (0.025)	0.153 (0.013)	0.190 (0.013)
Unemployed ($x_{19=1}$)	0.290 (0.017)	0.640 (0.011)	0.909 (0.012)	0.848 (0.026)	0.108 (0.008)	0.263 (0.013)
Public assist. ($x_{20=1}$)	0.239 (0.019)	0.729 (0.011)	0.967 (0.006)	0.839 (0.025)	0.083 (0.007)	0.202 (0.013)
Poverty ($x_{21=1}$)	0.230 (0.023)	0.810 (0.010)	0.970 (0.007)	0.819 (0.026)	0.015 (0.003)	0.200 (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
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 respective characteristic 'x', the highest conditional probability of a latent class is shaded gray. For example if density is above the national median, latent class 4 is most likely (probability is 0.850).

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

Note: goodness of fit statistics: BIC=1056817.983; entropy=0.945.