Chapter 8
What If Neighbors' Neighborhoods Differ? The Influence of Neighborhood Definitions of Health Outcomes in Accra

Alex Zvoleff, Li An, Justin Stoler, and John R. Weeks

Neighborhood context is recognized as an important predictor of individual-level behaviors and health outcomes (Pickett and Pearl 2001; Lee and Cubbin 2002; Sampson 2003). Neighborhoods, however, are difficult to define both in theory and in practice, and are often drawn to follow existing administrative boundaries or sampling schemes, or must be set arbitrarily due to a lack of sufficient data. Given the role of neighborhood context in influencing health outcomes, it is crucial that the area of influence surrounding the unit of analysis (be it a person, household, etc.) be properly defined. As already discussed in previous chapters, if we do not identify neighborhoods correctly, we cannot properly evaluate neighborhood effects. Defining neighborhoods is a challenge across the social sciences; investigations of the role of neighborhood context in decision making and shaping of individual-level outcomes are seen in public health, geography, demography, and sociology with no consistent approach to identifying and evaluating neighborhood effects. In this chapter, we outline several types of neighborhood definitions from the literature, and then, using data from the Women's Health Study of Accra, implement a spatial model together with a simulation approach to examine how two alternative neighborhood definitions affect modeling of individual-level health outcomes.

A. Zvoleff (✉) • L. An
Department of Geography, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182-4493, USA
e-mail: azvoleff@mail.sdsu.edu; lan@mail.sdsu.edu

J. Stoler
Department of Geography and Regional Studies, Department of Public Health Sciences, University of Miami, 1000 Memorial Drive, Coral Gables, FL 33124-2221, USA
e-mail: stoler@miami.edu

J.R. Weeks
Department of Geography, International Population Center, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182-4493, USA
e-mail: john.weeks@sdsu.edu

J.R. Weeks et al. (eds.), *Spatial Inequalities: Health, Poverty, and Place in Accra, Ghana*, GeoJournal Library 110, DOI 10.1007/978-94-007-6732-4_8,

To ground this work, we first discuss how neighborhoods have traditionally been represented in the literature. Neighborhood context has been recognized as an important component of models of many demographic and health indicators (e.g. fertility, morbidity, mortality), and is often accounted for by using multi-level models that account for hierarchically structured study designs (Subramanian 2010). Neighborhood effects represent the portion of the variance in an outcome variable that is attributable to individuals' shared experience as residents of a single area. Neighborhood effects account for common exposure to institutions, social networks, environments, etc. A key decision to be made in structuring models that account for neighborhood effects is the definition of neighborhoods. To ground our study, we briefly review several approaches to defining neighborhoods that have been used in the social science literature.

We distinguish between two major types of neighborhood definitions: a territorial definition and an egocentric definition. In a territorial definition, an individual belongs to a shared neighborhood based on a higher-level variable (such as administrative boundaries), or some other defining characteristic (local tradition, spatial homogeneity, etc.). A territorial neighborhood is shared by a set of individuals, all of whom reside within the same spatially contiguous neighborhood. By contrast, in an egocentric definition each individual is assigned to a unique neighborhood based on his or her location in space and, as implemented here, a fixed radius around that point. As noted in Chap. 3, both the scale and shape of neighborhoods can have important effects on modeling of health outcomes. It is important that analysts make informed choices regarding the tradeoffs between alternative approaches to parameterizing neighborhood effects.

One alternative to administrative boundary-defined neighborhoods is to construct divisions of space based on other criteria, such as social connections and the likely spatial background area inhabited by each individual. "Vernacular" neighborhoods (a territorial type of neighborhood definition) are perhaps the most intuitive approach. Vernacular neighborhoods are defined using local knowledge to construct a set of neighborhoods based on residents' commonly accepted divisions of an area, and are closest to the colloquial definition of neighborhood. Local residents might use a vernacular neighborhood name to, for example, direct a taxi to a particular area of a city (Weeks et al. 2010).

Vernacular neighborhood definitions are advantageous compared to exogenous constructions like census tract boundaries because they capture social dimensions not often considered in the design of survey sampling frames. As described in Chap. 2, vernacular neighborhoods are often defined based on a combination of census data, in-depth field work, and focus group research, which allows researchers to understand how residents define core and boundary regions of their neighborhoods. Once the boundaries are defined, if the sampling units of a relevant social survey data set are available at a finer scale than the final vernacular neighborhoods, construction of a spatial vernacular neighborhood-linked data set (including any number of other sources of spatial data) becomes a matter of agglomeration of survey units into the proper sets. When neighborhood boundaries and sampling unit boundaries do not coincide, as is often the case, vernacular neighborhood boundaries must be approximated, either by using the judgment of researchers to merge them into the most appropriate units, or by disaggregating them to a finer scale so that they can be appropriately recombined.

Another technique of constructing territorial neighborhoods is to use an "organic" approach. This method has been successfully used in Accra as a basis for mapping units determining contextual influences on processes including marriage timing and fertility (Weeks et al. 2010). Organic neighborhoods enjoy the advantage of having well-defined (by design) statistical properties. The AMOEBA (A Multidirectional Optimum Ecotope-Based Algorithm) approach works by using a local spatial autocorrelation statistic to define clusters of related spatial units using an iterative process (Aldstadt and Getis 2006). AMOEBA "let the data speak for themselves" by arriving at a set of empirically defined neighborhoods based on any characteristic chosen by the analyst (e.g. land cover, a demographic indicator, or some composite index). The organic approach allows maximization of the homogeneity of neighborhoods, while maximizing the heterogeneity between neighborhoods (Weeks et al. 2010), thereby closely matching our informal definition of a neighborhood as an area defined by its unique characteristics.

Whereas past approaches to neighborhood definition were limited to merging or splitting of areal units (as applied to a survey sampling design, for example) individual-level survey data are increasingly available in a geo-referenced format that offers the ability to construct new measures of neighborhood context defined at the individual or *egocentric* level. Individual-level measures of context allow us the ability to use individual-level spatially explicit modeling strategies, in addition to the more common multi-level modeling approach used to account for nested survey data.

Egocentric definitions are one way of operationalizing this conception of an individual-level measure of neighborhood effects (Reardon et al. 2008). Egocentric neighborhoods allow an explicit examination of the effects of the scale of analysis on measures of neighborhood effects. With an egocentric definition, alternative spatial scales can be tested and examined, and theoretical conceptions of scale can be compared to what is observed empirically (Lee et al. 2008). One limitation of the egocentric approach, including our implementation here, is that, similar to territorial approaches, it assumes that all areas of the egocentric neighborhood are equally accessible and influential. This problem can be addressed to some extent using inverse distance or other weighting schemes; however it can be difficult to define resistance layers for physical neighborhood effects at multiple scales such as roads, rivers, or compound walls.

One advantage of territorial neighborhood definitions over egocentric definitions is that for some applications (such as planning interventions or designing survey sampling schemes) broader-scale patterns are most important. One example is the broad-scaled mapping of health trends, as might be done from remotely sensed imagery (Weeks et al. 2007). As discussed in Chap. 4, in cases where geospatial data are sparse, the ability to identify boundary polygons for neighborhoods using remote sensing could be an important tool.

To explore the impact of alternative neighborhood definitions on our ability to predict individual-level health outcomes, we compare egocentric and territorial definitions using a spatial regression model and simulation approach. We first use spatial modeling to estimate neighborhood effects using each definition, and then use the results of this model as input to a simulation model of three representative areas of varying socioeconomic status in Accra, Ghana. This approach lets us take advantage of the precise estimates of coefficient values and standard errors that can be obtained with the spatial regression models, while also investigating the impact of each neighborhood definition type on modeling of health outcomes through time.

In the simulation model, we vary neighborhood context over time and compare how the egocentric versus territorial definitions predict the combined physical functioning score from the SF-36 survey questions, using real survey data for comparison. We hypothesize that egocentric definitions of neighborhood context will better represent the influence of neighborhood context on health outcomes than territorial definitions due to their ability to incorporate local neighborhood context without suffering from boundary effects typical of territorial definitions.

8.1 Analysis Part 1: Spatial Regression Approach

To test our hypothesis, we use data from the Women's Health Study of Accra (WHSA). The WHSA is a community-based longitudinal study focusing on a sample of women from the Accra Metropolitan Area (Hill et al. 2007). Initially, 200 enumeration areas (EAs, similar to census tracts) within Accra were randomly selected with probability proportionate to size (number of people). After creating a list of eligible women in these EAs, a sample of 3,200 women was drawn with probabilities fixed by the age group of each woman, and the socioeconomic status of her EA. See Hill et al. (2007) and Douptcheva et al. (2011) for details on the study design. The WHSA was first conducted in 2003 (WHSA-I), and a second wave was completed in 2008–2009 (WHSA-II).

Our outcome variable, physical functioning (PF) score, is a measure derived from a series of questions borrowed from the Short Form-36 (SF-36) questionnaire (Ware et al. 1994) administered during the WHSA. The SF-36 questionnaire is a short survey designed to be easily administered while providing reliable and consistent measures of general health. The SF-36 has proven to be a reliable means of measuring health through self-reports, and has been administered in numerous countries, in varying contexts. The SF-36 survey is equivalent to the RAND 36-Item Health Survey, with the exception of some scoring differences that are not relevant for calculation of the PF scores we discuss here (see Hays et al. 1993 for scoring details).

The PF score is an individual measure based on a series of ten questions that focus on the degree to which health limits a number of common physical activities such as walking, bending, and climbing stairs. The three possible responses to each question indicate the degree to which health limits an individual's ability to engage in each activity: yes – limited a lot, yes – limited a little, and no – not limited at all. Following the established methodology for calculating the PF score, the response to each of the ten items was recoded to a 0–100 scale (100 = not limited at all, 50 = limited a little, 0 = limited a lot), and the mean response calculated (Hays et al. 1993; Ware et al. 1994). With the PF score, as with SF-36 and RAND summary measures in general, a higher score indicates a better degree of health (meaning less limitation on physical functioning in the case of the PF score).

We choose the PF score for our study because previous work has shown physical functioning to be, of the eight SF-36 scales, the most highly correlated with physical health. A physical health summary measure (also taking into account bodily pain, general health, and limitations in work and daily activities) can also be derived from the SF-36 (Ware et al. 1994). For clarity of presentation and interpretation, we use the physical function score, which focuses on a more limited subset of the SF-36 questionnaire, for this study.

We use the percent of vegetated land cover (derived from high-resolution satellite imagery) as the measure of neighborhood context in our models. Vegetative land cover has been linked to neighborhood structure in Accra (Weeks et al. 2007; Stoler et al. 2012). Areas with a high percentage of vegetative cover tend to be city parks, higher income areas, or areas neighboring forests. Areas with low vegetative cover tend to be heavily built up areas, such as large industry or low income residential areas. At a fine scale, less vegetative cover can also be due to the influence of large roads. The lack of vegetation in low-income residential areas allows us to use vegetative cover as a proxy for neighborhood effects on PF score. We would expect to find lower PF scores in lower income areas (with less vegetative cover), due to reduced access to health services, poorer sanitation facilities, and increased likelihood of exposure to environmental hazards.

To map vegetative cover, we use two QuickBird multispectral images from April 2002 and January 2010 (2.4 m spatial resolution). Following radiometric correction from digital numbers to spectral radiance, the images were atmospherically corrected using the empirical line method, clouds were masked, and a vegetation/non-vegetation map was produced using NDVI thresholds. The final product offers a 2.4 m resolution map of vegetative cover for the Accra Metropolitan Area, as described in Chap. 4.

We define our neighborhood effects measure for two neighborhood definitions (territorial and egocentric). For our territorial neighborhoods, we calculate percent vegetative cover within "Field Modified Vernacular neighborhoods" (FMVs) that have been previously defined for Accra. The FMVs were constructed by modifying a map from the Ghana Statistical Service, following extensive fieldwork and focus group interviews, to create a set of 108 FMVs (see Weeks et al. 2012; Engstrom et al. 2011; and Chap. 2). For comparison with the territorial FMV approach, we use an egocentric approach where we calculate the percent of vegetative cover within a circular buffer of each woman's household (obtained via GPS during the WHSA-II survey). Because our *percent cover* measure is derived from remote sensing imagery, there are some missing values in the data (areas of cloud cover, or where a portion of a woman's egocentric neighborhood extends off the image). In some cases, a large portion of the land area surrounding a woman's location is unobserved in our imagery; this occurs for women whose location is obscured by clouds in the image, and for women located near the edge of the study site. To correct for this problem, we calculate the percent cover of each egocentric neighborhood using the total observed neighborhood area in the denominator, rather than total neighborhood area (as would normally be done for a percentage calculation). This presumes that unobserved areas of a particular egocentric neighborhood have the same average land cover composition as the observed area of that same neighborhood. To minimize bias from incomplete observations, we drop from our analysis all women where data are missing for more than 25 % of the land cover pixels in their egocentric neighborhood. This yields a final sample size of 1,114 women from WHSA-I, and 2,279 women from WHSA-II.

To determine the size (radius) of the egocentric neighborhood for our models, we examine the relationship between the egocentric variable (log percent vegetation) and PF score at a range of radii. We find that the relationship levels off with increasing radius (remembering that the area of each egocentric neighborhood increases with the square of the radius). Around about 700 m radius, the correlation coefficient between percent vegetation and PF score levels off at approximately 0.05 and is statistically significant ($p < 0.05$). We therefore chose to use a 700 m radius for our models. Though it is possible we would see changes beyond a 1,000 m radius, it is unlikely that the results for these egocentric neighborhoods would differ from those for the FMV neighborhoods we present here (unless constructed with far larger radii than we would consider reasonable, in which case we would expect the correlation to converge to zero).

We present two regression models to compare two neighborhood contexts (FMV and egocentric). In both cases, the metric is the log of percent vegetative cover. We use the log transformation of percent cover to improve the distribution of the otherwise highly positively skewed data. The log transformation also conforms to our expectation, and findings from other studies, such as Yabiku (2006), that the difference between neighborhoods with low levels of vegetative cover (e.g., 20 % compared to 10 %) is likely to have more of an impact on an individual than the difference at high levels (e.g., between 90 and 80 %). Even though the absolute change is the same, a change from 20 to 10 % is likely to be more noticeable to an individual, as it represents a halving of the vegetative cover in a neighborhood.

We first experimented with using ordinary least squares (OLS) regression models for our models of PF scores. However, given the nature of our data, we expected that spatial autocorrelation might make an OLS approach inappropriate. Spatial autocorrelation may exist in the residuals due to the presence of unobserved covariates (indicating a spatial error model might be appropriate), or due to spatial dependence in the dependent variable itself (suggesting a spatial lag model), or due to both simultaneously (Anselin and Lozano-Garcia 2009; Getis 2009). Lagrange multiplier (LM) tests (Anselin and Rey 1991) were used to test for spatial effects in the residuals of the ordinary least-squares (OLS) representations of both the egocentric and FMV models presented in Table 9.1. In both cases, the LM tests indicated statistically significant spatial lag effects.

Failure to account for spatial autocorrelation can lead to bias and inefficient estimates of regression coefficients (Getis 2009). To account for this spatial autocorrelation, we use a spatial simultaneous autoregressive lag model. In addition to the usual data matrix (X), coefficient matrix (β), and error term (ε) of an ordinary least-squares (OLS) regression, spatial lag models include an additional term with spatially lagged values of the dependent variable (y), weighted according to a weights matrix (W) and autoregressive lag coefficient (ρ) (LeSage and Pace 2009). This results in the model formula:

$$y = \rho W y + X\beta + \epsilon$$

In our models, we define the spatial weights matrix as including as neighbors all surveyed households within 700 m of each individual, using inverse distance weighting. The inverse distance weighting reflects our expectation that neighboring households will be more closely related to each other than to distant households. All the models we present here were calculated in R version 2.14.2 (R Development Core Team 2012) using version 0.5–45 of the 'spdep' package (Bivand 2012). A multilevel model is another potential approach, particularly for the FMV neighborhood model; we use a spatial lag model for both models to simplify intercomparisons. The inclusion of spatial effects within a multilevel model is an active research area (Corrado and Fingleton 2011).

The results of the two spatial regressions testing the influence of the two different parameterizations of neighborhood-level percent vegetative cover on predicting PF scores are in Table 8.1. Both models include the same controls; however, the egocentric model includes a measure of percent vegetative cover calculated over a 700 m egocentric neighborhood, while the FMV model uses a measure of vegetative cover calculated over a territorial neighborhood. For both models, the LM test p-values indicate no significant spatial autocorrelation in the residuals due to spatial lag.

The coefficients for age and age-squared are both highly significant, and in the expected directions in both models. While the linear term on age indicates higher PF scores with increasing age, this effect is attenuated by the negative coefficient on the quadratic term. The Ga are associated with higher PF scores than all other ethnic groups (although the only marginally significant effect is between the Ga and Akan, $p < 0.10$). As expected, education is positively associated with PF, with greater PF scores as the level of education increases ($p < 0.05$ for all the terms except for the difference between the no-schooling and primary schooling-only groups).

The relationship between the use of charcoal in cooking and PF score is not clear; the effect is surprisingly positive, although with a large standard error. We would expect charcoal usage to be negatively associated with PF scores, as charcoal usage was intended to act as a proxy for lower income households, who likely have less access to health services. Furthermore, there are direct effects of charcoal on health, which makes the positive coefficient surprising, though not significant. Ownership of a toilet, an indicator of higher socioeconomic status, and possibly of less exposure to environmental health hazards, is also unexpectedly negatively associated with PF

Table 8.1 Comparison of egocentric neighborhoods and field-modified vernacular neighborhoods (FMV) spatial lag models

	700 m radius egocentric neighborhood model				Field-modified vernacular neighborhood (FMV) model			
	Beta	Std. Err.	Prob. (> |z|)		Beta	Std. Err.	Prob. (> |z|)	
log(percent vegetation egocentric)	1.50	0.78	0.055	·				
log(percent vegetation FMV)					1.86	0.69	0.007	**
Age	0.34	0.13	0.007	**	0.34	0.13	0.007	**
Age^2	−0.01	0.00	< 0.001	***	−0.01	0.00	< 0.001	***
Ethnicity (Akan)	−1.70	0.96	0.076	·	−1.68	0.96	0.080	·
Ethnicity (Ewe)	−1.34	1.24	0.278		−1.26	1.24	0.308	
Ethnicity (Other)	−0.96	1.26	0.444		−0.75	1.26	0.554	
Education (primary)	2.38	1.50	0.113		2.46	1.50	0.100	
Education (middle)	3.42	1.21	0.005	**	3.45	1.21	0.004	**
Education (secondary)	3.76	1.48	0.011	*	3.79	1.48	0.010	*
Education (higher education)	4.08	1.71	0.017	*	4.04	1.71	0.018	*
Charcoal for cooking	0.30	0.93	0.750		0.41	0.94	0.663	
Has own toilet	−1.53	0.95	0.109		−1.63	0.95	0.087	·
Years in house (5–15 years)	0.47	1.20	0.694		0.36	1.20	0.762	
Years in house (15–30 years)	−0.53	1.20	0.661		−0.64	1.20	0.596	
Years in house (>30 years)	−2.92	1.37	0.032	*	−3.02	1.36	0.027	*
Intercept	88.44	3.93	< 0.001	***	87.57	3.82	< 0.001	***
Rho		0.016				0.016		
Rho standard error		0.008				0.008		
Log likelihood		−9,878.99				−9,877.25		
AIC		19,793.98				19,790.50		
n		2,279				2,279		
LM test statistic		0.187				0.162		
LM test p-value		0.665				0.687		

Note: Both models use spatial weights matrix of neighbors within a 700 m radius, with weights inversely proportional to distance. All factors are coded with treatment contrasts: for ethnicity coefficients, Ga is the reference class; for education, "no education" is the reference class; for "Years in house", 0–5 years living in the house is the reference class

Significance is coded as: · for $p < .1$; * for $p < .05$; ** for $p < .01$; *** for $p < .001$

score (insignificant in the egocentric model, $p < 0.10$ level in FMV model). The time of residence in the area is significant only for the coefficient for residents who have lived in their home for more than 30 years (who have significantly lower PF scores than those who have been resident for 5 years or less).

The log percent vegetation term is highly significant for the FMV model ($p = 0.007$), and marginally so for the egocentric model ($p = 0.055$). The term is positive in both models as expected: areas with higher vegetative cover tend to be of higher socio-economic status. Residents in lower socioeconomic status and industrial areas are expected to have higher exposure to environmental stressors, and to have worse access to sanitation and health services (see Chap. 7), thereby resulting in lower PF scores. The size of the neighborhood effect increases and is more significant with the FMV term than with the egocentric term, although the standard errors for the parameters overlap. Judging by log likelihood and by AIC, the FMV model provides a better fit to the data.

8.2 Analysis Part 2: Spatial Simulation Approach

With the completed spatial regressions, we explore how an egocentric approach and territorial approach compare in predicting PF scores over time using a simulation approach. The results of the spatial regression models are used to predict the physical functioning scores for a sample of WHSA respondents from three representative areas of Accra. Agent-based (or individual-based) models have seen increasing usage in the geographic literature for modeling individual-level decisions and interactions in non-linear systems (Epstein 1999; Axtell et al. 2002; Parker et al. 2003; Grimm et al. 2005; An et al. 2005). Agent-based simulation models are advantageous for study of these systems as they can offer a sort of laboratory in which alternative hypotheses can be tested. Here we use a simple simulation approach to model physical functioning scores of a set of women from each of three different neighborhoods, using data from the WHSA-I to test a series of regression models derived from the WHSA-II surveys. The simulation modeling framework has the advantage of allowing us to test, over time, how the two alternative neighborhood definitions compare in their ability to model neighborhood contextual influences on health outcomes. While the spatial regression approach allows us to estimate effect sizes and significance, it is difficult to visualize the implications of these findings on a broader scale. By using simulation modeling together with spatial regression, we leverage the advantages of each of these two spatial modeling approaches, to more fully test our hypothesis that egocentric neighborhoods will serve as better indicators of neighborhood contextual effects than will territorial definitions.

We chose three FMV neighborhoods for this analysis: Nima, North Kaneshie, and Dansoman Estates (see Table 8.2 and Fig. 8.1). Nima is a well-known and densely populated neighborhood of low socio-economic status. North Kaneshie, located north-west of Nima, represents a middle-income area with slightly lower

Table 8.2 Comparison of characteristics of the three neighborhoods (using data from the WHSA-II)

Neighborhood	Mean PF score	Mean SES	Mean education	Mean age	n
Nima	83.96	1.46	2.27	45.59	74
North Kaneshie	79.68	2.95	3.00	48.86	45
Dansoman Estate	82.26	3.87	3.47	45.50	51

Note: The sample size (n) listed here is the sample size for the total number of women in that neighborhood from the WHSA-II survey. The ABM sample is smaller (see text)

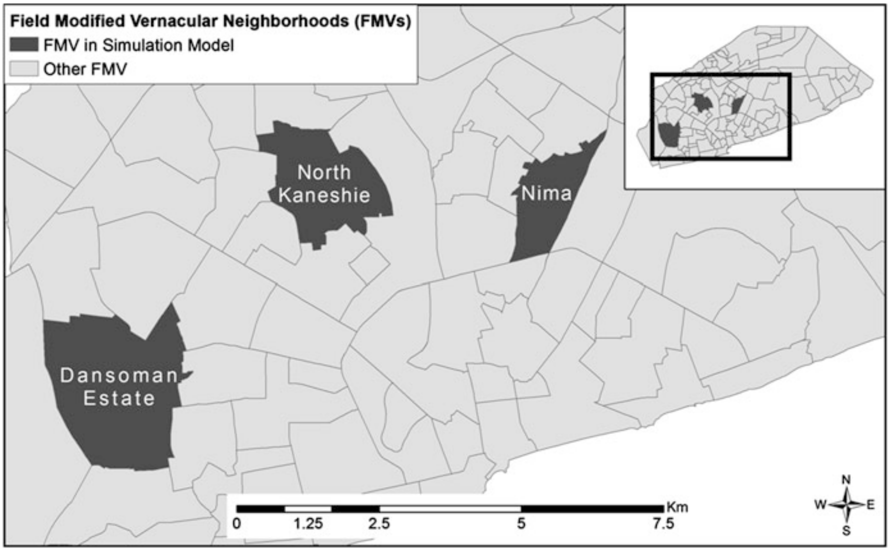

Fig. 8.1 The three representative areas of the Accra metropolitan area included in the simulation model. The three areas coincide with field-modified vernacular (FMV) neighborhoods

PF scores on average than Nima or Dansoman Estates, slightly older residents on average, and moderate educational levels. Dansoman Estates is a higher income area with higher education levels but with PF scores and mean ages similar to Nima. For the simulation model, we use sets of residents from each of the three areas who lived in the same location during both the WHSA-I and WHSA-II surveys. By using these three areas for comparison, we are able to focus in detail on the processes occurring at a neighborhood level. In addition, since the enumeration areas sampled during the WHSA are not spatially contiguous, our analysis of each of these three areas independently also simplifies interpretation of our simulation modeling results.

To run the simulation model, we need to model the change in neighborhood context over time. In the simple case presented here, we model land cover change to predict the percent of vegetated land cover in each neighborhood at each time step. There are many models of land cover change of varying complexity in the literature (Brown et al. 2000; Lambin et al. 2001; Petit et al. 2001; An et al. 2006; Millington

Table 8.3 Markov transition matrices for the three Nima clusters (rows are from 2002 image, columns are from 2010 image)

	Non-vegetation	Vegetation
Nima cluster		
Non-vegetation	0.9931	0.0069
Vegetation	0.0069	0.9931
North Kaneshie cluster		
Non-vegetation	0.9868	0.0132
Vegetation	0.0244	0.9756
Dansoman Estates cluster		
Non-vegetation	0.985	0.015
Vegetation	0.0244	0.9756

et al. 2007; An and Brown 2008). As our primary concern is to understand the effects of alternative neighborhood definitions on models of physical functioning rather than to model the generative process behind land-cover change in Accra *per se*, we choose a series of simple first-order stationary Markov models of land-cover change as inputs to our simulation model. We generate one Markov model for each of the three representative areas we consider, using the two classes *vegetation* and *non-vegetation*.

A Markov model is a simple model of land change that represents change using a matrix of transition probabilities. Markov models rely on the assumption that land-cover change can be represented as a stochastic process in which the probable future state of a system can be derived based on knowledge of prior states of the system. Using a transition matrix, a Markov chain can be constructed, representing transition of the landscape between a series of states, with each state dependent on only the transition probabilities, and the last observed (or modeled) state of the system (Brown et al. 2000; Petit et al. 2001).

As we have only two observations (2002 and 2010), we assume a constant rate of change in each neighborhood in the 8-year period between the two images. In a typical Markov model, the probability of transition between states is calculated by considering the probability of transition at a pixel level, by considering the initial and final states of each pixel. However, due to misregistration and misclassification, comparison of two images at the pixel-level can lead to high levels of error, and inaccurate transition probabilities. To alleviate this problem, we calculate the transition probabilities used in the model over a five-by-five pixel moving window (Brown et al. 2000). After calculating the transition probabilities using the land cover map from the 2002 and 2010 QuickBird images, we use matrix algebra to correct the transition probabilities (derived from an 8-year window) to apply to the 1-year time step on which we run the model (Petit et al. 2001). After this calculation, we obtain the final transition matrices (Table 8.3).

Though we would prefer to parameterize our simulation model using regressions from the WHSA-I dataset, GPS coordinates were not collected for WHSA-I respondents, so we are unable to accurately calculate egocentric neighborhoods for these women. Using the WHSA-II dataset to parameterize our simulation model is, in a sense, cheating, as we are using future data to parameterize a model that is then run starting in the past. As we do have GPS coordinates for women from WHSA-II, we attempted to run our regressions on the WHSA-I panel, by only using women known to have resided in the same location in WHSA-II and I (so that we have precise location data for egocentric neighborhood calculation, by using matched coordinates from the 2008 WHSA-II survey). This restriction drastically reduces the sample size, limiting our ability to obtain unbiased estimates from the data. Therefore, we use WHSA-II data to parameterize our models, and run the simulations starting in the year 2003, assigning characteristics to the women in the simulation directly from the WHSA-I data. For the simulation models, we only use women surveyed in both WHSA-I and II. This ensures we can compare the predicted PF scores from 2008 from the simulation models with the observed scores from the same women from the WHSA-II survey conducted in 2008.

We run the model using three different types of scenarios. The first scenario includes no neighborhood effect (using the regression coefficients from a spatial lag model without a neighborhood effects term, but with identical controls to the models presented in Table 8.1). The second type of scenario uses the results of the egocentric spatial lag model presented earlier, with percent land-cover calculated for each woman at the beginning of each time step over a 700 m egocentric neighborhood. The third type of scenario uses the FMV neighborhood definition, with percent land cover calculated for each woman according to the FMV neighborhood in which she resides. We integrate our Markov models of land cover change to provide the land cover estimates needed in these models. For each of the three representative areas (Nima, North Kaneshie, and Dansoman Estates), we calculate independent Markov transition matrices, allowing the pace of land cover change to vary for each area.

For each scenario, the model runs on a 1-year time step. The model is initialized with the set of WHSA women resident in the neighborhood during WHSA-I, and land cover is assigned to the neighborhood as a grid of 2.4 m resolution cells (as a binary vegetation/non-vegetation map) based on a 2003 QuickBird classification. At the beginning of each time step, the PF score is predicted for each woman using the appropriate model from Table 8.1, and the appropriate neighborhood definition to calculate percent vegetative cover depending on whether it is an egocentric, territorial (FMV) or a no-neighborhood-effects scenario. Land cover change is then modeled using the appropriate Markov transition matrix for the neighborhood, and the resulting land cover is carried forward to the next time step. At the end of each time step, the age and length of time at her residence are incremented for each woman.

For each of the three test areas (Nima, North Kaneshie, and Dansoman Estates), we run each of the three types of scenarios, resulting in a total of nine different simulations. For each simulation, we ran 35 model runs, a sufficient number of model runs to ensure a good estimate of the variance in the mean of the model runs (the variance did not increase as we added additional runs). We present the mean of these 35 model runs in our results to show the spread of variability that manifests from the stochasticity in the land-change model. Given the low amount of stochasticity in these simulations, the variance around these means is essentially invisible. This is not meant to overstate the accuracy of our predictions, but rather to indicate *a priori* that the model runs, by design, do not have a high degree of variability in their results within each of the nine different simulations.

We run each simulation for 17 time steps, representing the years 2003–2020. The initial time step therefore coincides with the WHSA-I survey date, and the sixth time step (2008) coincides with the WHSA-II survey date. Mortality is not included in these models. We experimented with including age dependent mortality in the model using World Health Organization (WHO) data; however, age-specific mortality is not available with a high enough degree of temporal resolution for Accra (WHO data are only available in three age groups: 0–15 years, 15–60 years, and over 60 years old). Neglecting mortality is not an issue for the WHSA-I to WHSA-II time interval (2003–2008) that we focus on in this paper, as the women included in the model represent real women who were surveyed and known to be alive in both 2003 and 2008. We continue our model runs after 2008 not as a predictive tool, but to descriptively compare the evolution of the PF scores as simulated on a longer timescale. If we were focused on population dynamics or on making accurate predictions of PF for these neighborhoods out to 2020, neglecting mortality would not be defensible.

The simulation results for the three representative neighborhoods are presented in Fig. 8.2. The most noticeable trend is the decline in PF scores over time. This trend is due to the increasing age of the respondents in our cohort. As we restrict our sample to include only women sampled in the same location in both the WHSA-I and WHSA-II surveys, the average age of our sample is relatively high (around 48 years old at the beginning of the model). This is in part because younger, more mobile women have been filtered out (as they would not be in the same household twice). The decline in mean PF score can also be seen to increase with time; this is due to the effect of the quadratic term on age.

We can also compare the three models across each scenario. In all three areas, the mean PF score as modeled by the no neighborhood effects model, egocentric model, and FMV (vernacular) neighborhoods models are all very similar. This is due to the relatively small effect of neighborhood context (as determined by the regression models) on PF score as compared to the size of the age effect. The age effect dominates when we model PF scores through time. The FMV model tends to predict higher PF scores than the egocentric model, due to the influence of the larger intercept term in the FMV model. Through time, the FMV neighborhood and egocentric neighborhood models track closely together, although the no-neighborhood-effects model diverges slightly, most evidently in the North Kaneshie simulation. Here we see the small, though visible, impact of ignoring neighborhood effects in our model.

In Table 8.4 we compare the mean PF scores for each area as calculated by the simulation models using the PF scores from the survey data. We see the closest match between our model output and the observed data for the Nima and North Kaneshie neighborhoods. In both cases the FMV neighborhood definition tracks a bit more closely with the observed data. However, in North Kaneshie the observed data (as seen in the left two columns of Table 8.4) shows an increase in the

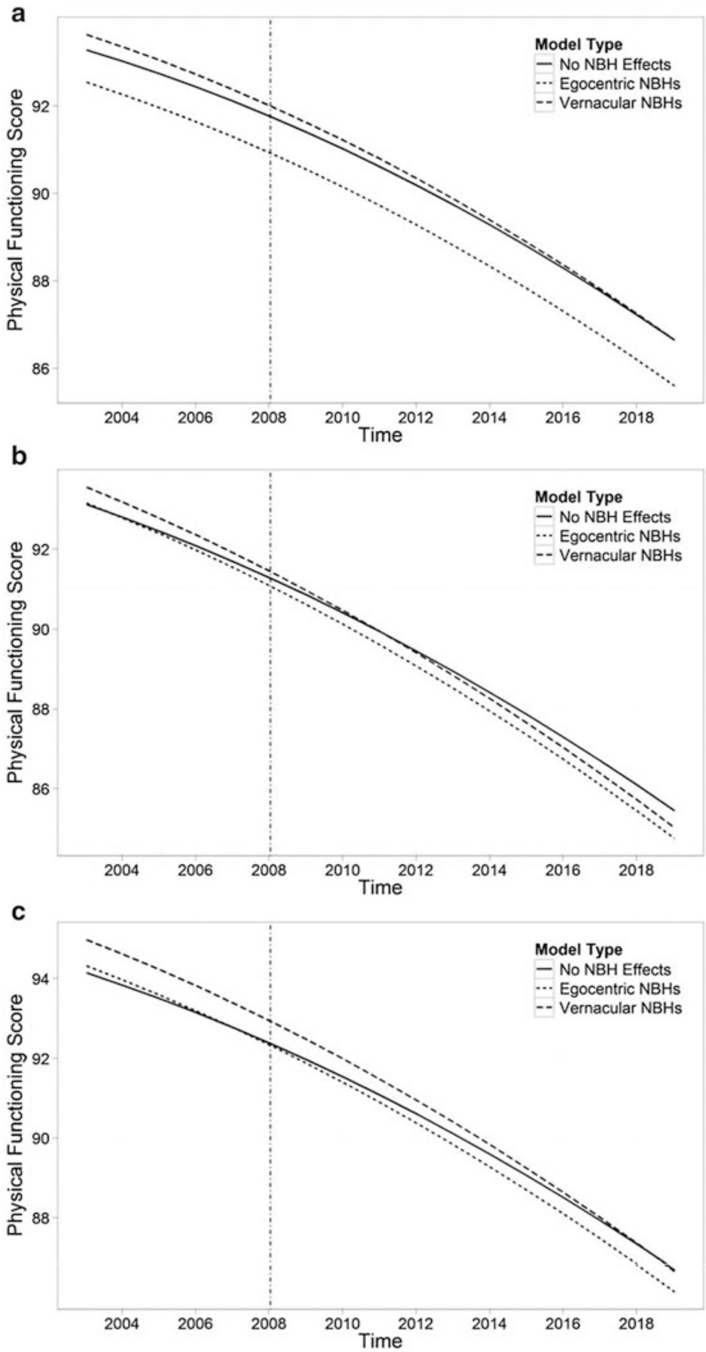

Fig. 8.2 The simulation results for (**a**) Nima, (**b**) North Kaneshie and (**c**) Dansoman estates. The *vertical dotted line* in each figure represents WHSA-II survey date

Table 8.4 Comparison of the mean physical functioning (PF) score of the three neighborhoods included in the simulation model, from the WHSA surveys and from the model results

Neighborhood	WHSA-I (2003)	WHSA-II (2008–2009)	Modeled 2009 (Egocentric)	Modeled 2009 (FMV)
Nima	95.94	91.89	90.92	91.99
North Kaneshie	89.53	92.78	91.07	91.43
Dansoman Estate	87.64	87.65	92.32	92.93

Note: Column 2 is the mean PF score from the WHSA-1 survey. Column 3 is the mean PF score from the WHSA-2 survey. Column 4 is the mean PF score from 2009 from the simulation, modeled using the egocentric neighborhood definition. Column 5 is the mean PF score from 2009 from the simulation, modeled using the FMV neighborhood definition

mean PF score; our model greatly over-predicted the 2003 observed values and produced a decrease in PF score up to 2008 that ended up matching the observed data. Clearly the model requires fine-tuning if we were to seek to match the process more closely in North Kaneshie. In Dansoman estates we see essentially no change in the observed mean PF scores, but the simulation model again predicts a decrease. Spatial heterogeneity in the observed relationship between PF score and neighborhood context is a possibility that should be further explored.

8.3 Conclusions

Although we expected to derive more predictive power from the egocentric neighborhood definition than from territorial measures, our results do not indicate that the egocentric approach is superior to the territorial FMV definition. The spatial regression models indicate that the size of the neighborhood effect from an egocentric neighborhood is similar (though less) in magnitude to that measured from a territorial neighborhood. To further compare the two neighborhood definitions and explore their dynamics over time, we used a simulation modeling approach. Our simulation results show a similar pattern. For the prediction of 2008 PF scores, the territorial and egocentric neighborhood definitions both perform similarly, and on a longer timescale, the results predicted under both definitions track closely together. The egocentric metric applied here, percent of vegetated land cover within a given distance buffer, is exceedingly simple. A more advanced approach that considered texture measures from the imagery, as demonstrated in Chap. 5, might better account for important neighborhood characteristics such as sharp boundaries between areas. Our comparison of the observed data with our simulation model results suggests that there may be spatial heterogeneities in the relationship between PF score and neighborhood context. Future work to untangle this relationship would inform our understanding of the relation between neighborhood context and health.

The egocentric neighborhood approach shows promise for estimating the relevant scale to be used in modeling neighborhood effects on individual-level health, and more sophisticated techniques of measuring these neighborhoods may unlock additional predictive power. Although we see similar effects between the egocentric and territorial definitions for PF score as our outcome variable, the same may not be true in general when other health outcomes are considered. An understanding of the spatial process itself is essential to choosing the proper parameterization of neighborhood context. When calculated as a continuously varying measure at the individual level, egocentric neighborhood context measures may be a useful addition to the analyst's tool kit. With the ability to better control for spatial autocorrelation in this type of modeling framework, the use of egocentric neighborhoods presents a potential alternative to multilevel modeling techniques traditionally used for exploring neighborhood effects.

Acknowledgments This research was funded in part by grant number R01 HD054906 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development ("Health, Poverty and Place in Accra, Ghana," John R. Weeks, Project Director/Principal Investigator). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Child Health and Human Development or the National Institutes of Health. Additional funding was provided by Hewlett/PRB ("Reproductive and Overall Health Outcomes and Their Economic Consequences for Households in Accra, Ghana," Allan G. Hill, Project Director/Principal Investigator). The 2003 Women's Health Study of Accra was funded by the World Health Organization, the US Agency for International Development, and the Fulbright New Century Scholars Award (Allan G. Hill, Principal Investigator).

References

Aldstadt, J., & Getis, A. (2006). Using AMOEBA to create a spatial weights matrix and identify spatial clusters. *Geographical Analysis, 38*(4), 327–343. Last accessed 6 Mar 2012.

An, L., & Brown, D. G. (2008). Survival analysis in land change science: Integrating with GIScience to address temporal complexities. *Annals of the Association of American Geographers, 98*(2), 323–344.

An, L., Linderman, M., Qi, J., Shortridge, A., et al. (2005). Exploring complexity in a human environment system: An agent-based spatial model for multidisciplinary and multiscale integration. *Annals of the Association of American Geographers, 95*(1), 54–79.

An, L., He, G., Liang, Z., & Liu, J. (2006). Impacts of demographic and socioeconomic factors on spatio-temporal dynamics of panda habitat. *Biodiversity and Conservation, 15*(8), 2343–2363. Last accessed 15 May 2010.

Anselin, L., & Lozano-Garcia, N. (2009). Spatial hedonic models. In T. C. Mills & K. D. Patterson (Eds.), *Palgrave handbook of econometrics*. Basingstoke/New York: Palgrave Macmillan. Last accessed 20 Jan 2012.

Anselin, L., & Rey, S. (1991). Properties of tests for spatial dependence in linear regression models. *Geographical Analysis, 23*(2), 112–131. Last accessed 30 Mar 2012.

Axtell, R. L., et al. (2002). Population growth and collapse in a multiagent model of the Kayenta Anasazi in Long House Valley. *Proceedings of the National Academy of Sciences, 99*(90003), 7275–7279.

Bivand, R. (2012). *spdep*: *Spatial dependence*: *Weighting schemes*, *statistics and models*. http://CRAN.R-project.org/package=spdep

Brown, D. G., Pijanowski, B. C., & Duh, J. D. (2000). Modeling the relationships between land use and land cover on private lands in the Upper Midwest, USA. *Journal of Environmental Management, 59*, 247–263.

Corrado, L., & Fingleton, B. (2011). *Multilevel modelling with spatial effects*. University of Strathclyde Business School, Department of Economics. http://ideas.repec.org/p/str/wpaper/1105.html

Douptcheva, N., Hill, A. G., Adanu, R. M. K., Anarfi, J., Blanchard, K., Fink, G., et al. (2011). *Final report on the women's health study of Accra, Wave II*. Accra: Institute for Statistical, Social and Economic Research, University of Ghana, Legon.

Engstrom, R., Ashcroft, E., Jewell, H., & Rain, D. (2011). Using remotely sensed data to map variability in health and wealth indicators in Accra, Ghana. In *Urban Remote Sensing Event (JURSE), 2011 Joint* (pp. 145–148), Munich.

Epstein, J. M. (1999). Agent-based computational models and generative social science. *Complexity, 4*(5), 41–60. Last accessed 16 Sept 2010.

Getis, A. (2009). Spatial autocorrelation. In M. M. Fischer & A. Getis (Eds.), *Handbook of applied spatial analysis: Software tools, methods and applications* (pp. 255–278). Berlin: Springer.

Grimm, V., et al. (2005). Pattern-oriented modeling of agent-based complex systems: Lessons from ecology. *Science, 310*(5750), 987–991.

Hays, R. D., et al. (1993). The RAND 36-item health survey 1.0. *Health Economics*, *Health Economics*. *2*(3), 217–227. Last accessed 16 Apr 2012.

Hill, A. G., et al. (2007). Health of urban Ghanaian women as identified by the Women's Health Study of Accra. *International Journal of Gynecology & Obstetrics, 99*(2), 150–156. Last accessed 3 Apr 2012.

Lambin, E. F., et al. (2001). The causes of land-use and land-cover change: Moving beyond the myths. *Global Environmental Change, 11*(4), 261–269. Last accessed 28 Jan 2010.

Lee, R. E., & Cubbin, C. (2002). Neighborhood context and youth cardiovascular health behaviors. *American Journal of Public Health, 92*(3), 428–436. Last accessed 3 May 2012.

Lee, B. A., et al. (2008). Beyond the census tract: Patterns and determinants of racial segregation at multiple geographic scales. *American Sociological Review, 73*(5), 766.

LeSage, J. P., & Pace, R. K. (2009). Spatial econometric models. In M. M. Fische & A. Getis (Eds.), *Handbook of applied spatial analysis: Software tools, methods and applications* (pp. 255–278). Berlin: Springer.

Millington, J., Perry, G., & Romero-Calcerrada, R. (2007). Regression techniques for examining land use/cover change: A case study of a Mediterranean landscape. *Ecosystems, 10*(4), 562–578.

Parker, D. C., Manson, S. M., Janssen, M. A., Hoffmann, M. J., et al. (2003). Multi-agent systems for the simulation of land-use and land-cover change: A review. *Annals of the Association of American Geographers, 93*(2), 314–337. Last accessed 27 Jan 2010.

Petit, C., Scudder, T., & Lambin, E. F. (2001). Quantifying processes of land-cover change by remote sensing: Resettlement and rapid-cover changes in south-eastern Zambia. *International Journal of Remote Sensing, 22*(17), 3435–3456.

Pickett, K. E., & Pearl, M. (2001). Multilevel analyses of neighbourhood socioeconomic context and health outcomes: A critical review. *Journal of Epidemiology and Community Health, 55*(2), 111–122.

R Development Core Team. (2012). *R*: *A language and environment for statistical computing*. Vienna, Austria. http://www.R-project.org/

Reardon, S. F., et al. (2008). The geographic scale of metropolitan racial segregation. *Demography, 45*(3), 489–514.

Sampson, R. J. (2003). The neighborhood context of well-being. *Perspectives in Biology and Medicine, 46*(3 Suppl), S53–S64. Last accessed 3 May 2012.

Stoler, J., et al. (2012). Assessing the utility of satellite imagery with differing spatial resolutions for deriving proxy measures of slum presence in Accra, Ghana. *GIScience and Remote Sensing, 49*(1), 31–52. Last accessed 3 May 2012.

Subramanian, S. V. (2010). Multilevel modeling. In M. M. Fischer & A. Getis (Eds.), *Handbook of applied spatial analysis* (pp. 507–525). Berlin/Heidelberg: Springer. http://dx.doi.org/10.1007/978-3-642-03647-7_24

Ware, J. E., Kosinski, M., & Keller, S. D. (1994). *SF-36 physical and mental health summary scales: A users manual*. Boston: The Health Institute, New England Medical Center.
Weeks, J. R., Hill, A. G., Stow, D., Getis, A., & Fugate, D. (2007). Can we spot a neighborhood from the air? Defining neighborhood structure in Accra, Ghana. *GeoJournal, 69*(1), 9–22.
Weeks, J. R., Getis, A., Hill, A. G., Agyei-Mensah, S., & Rain, D. (2010). Neighborhoods and fertility in Accra, Ghana: An AMOEBA-based approach. *Annals of the Association of American Geographers, 100*(3), 558–578.
Weeks, J. R., et al. (2012). Connecting the dots between health, poverty and place in Accra, Ghana. *Annals of the Association of American Geographers*. Last accessed 8 June 2012.
Yabiku, S. (2006). Land use and marriage timing in Nepal. *Population and Environment, 27*(5), 445–461.