



Capturing the two dimensions of residential segregation at the neighborhood level for health research

Masayoshi Oka^{1,2*} and David W. S. Wong^{3,4}

¹ Division of Public Health Sciences, Department of Surgery, School of Medicine, Washington University in St. Louis, St. Louis, MO, USA

² Division of Epidemiology and Public Health, School of Medicine, University of Alcalá, Alcalá de Henares, Spain

³ Department of Geography and Geoinformation Science, College of Science, George Mason University, Fairfax, VA, USA

⁴ Department of Geography, University of Hong Kong, Pokfulam, Hong Kong

Edited by:

Monica Christine Jackson, American University, USA

Reviewed by:

Mildred Audrey Pointer, North Carolina Central University, USA
Guoqing Hu, Central South University, China

*Correspondence:

Masayoshi Oka, Division of Public Health Sciences, Department of Surgery, School of Medicine, Washington University in St. Louis, 660 S. Euclid Avenue, Campus Box 8100, St. Louis, MO 63110, USA
e-mail: moka@post.harvard.edu

Two conceptual and methodological foundations of segregation studies are that (i) segregation involves more than one group, and (ii) segregation measures need to quantify how different population groups are distributed across space. Therefore, percentage of population belonging to a group is not an appropriate measure of segregation because it does not describe how populations are spread across different areal units or neighborhoods. In principle, evenness and isolation are the two distinct dimensions of segregation that capture the spatial patterns of population groups. To portray people's daily environment more accurately, segregation measures need to account for the spatial relationships between areal units and to reflect the situations at the neighborhood scale. For these reasons, the use of local spatial entropy-based diversity index (SH_i) and local spatial isolation index (S_i) to capture the evenness and isolation dimensions of segregation, respectively, are preferable. However, these two local spatial segregation indexes have rarely been incorporated into health research. Rather ineffective and insufficient segregation measures have been used in previous studies. Hence, this paper empirically demonstrates how the two measures can reflect the two distinct dimensions of segregation at the neighborhood level, and argues conceptually and set the stage for their future use to effectively and meaningfully examine the relationships between residential segregation and health.

Keywords: residential segregation, local spatial entropy-based diversity index, local spatial isolation index, racial/ethnicity segregation, socioeconomic segregation

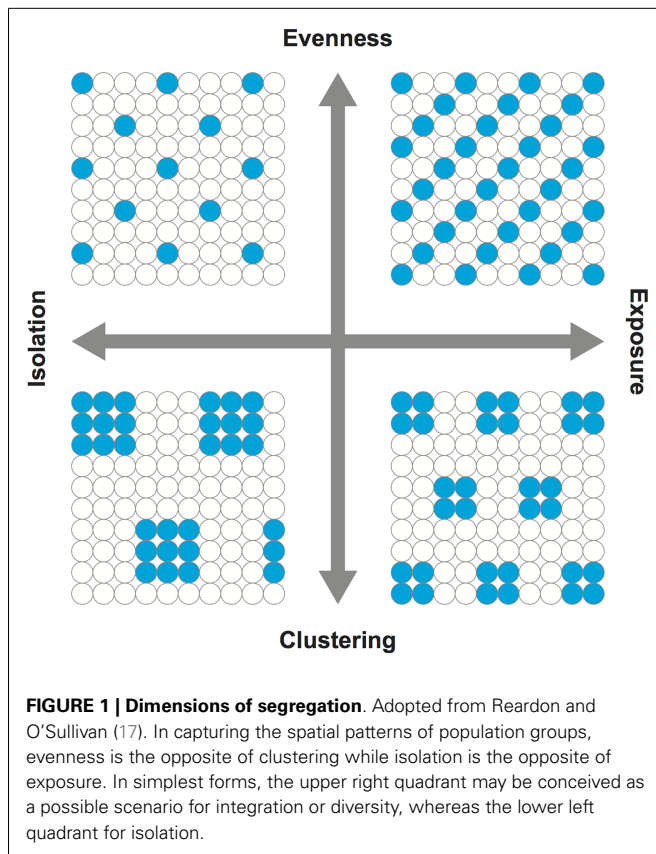
INTRODUCTION

In the United States (US), residential segregation has long been considered to shape health (i.e., health behaviors and health outcomes) as blacks and/or poor individuals are not distributed across geographic locations in the same manners as other groups (1–3). Note that all racial groups in this paper refer to non-Hispanic populations. A recent review article shows a growing number of US studies focusing on such concerns (4). However, the conceptual and methodological inconsistencies across previous studies limit the ability to draw specific conclusions about the relationships between residential segregation and health.

For instance, previous studies (4) have adopted either a global or local approach in measuring segregation. The former approach is based on the use of global measures that summarize the condition of a county or metropolitan area as a whole; these measures are used for inter-city, inter-regional, or inter-metropolitan comparisons. On the other hand, the latter approach is based on the use of local measures derived from data for US census tracts or block groups (sometimes zip-code areas); these measures are used for neighborhood comparisons. While these two approaches have different purposes, they also have different degrees of relevance to health research. Known as the modifiable areal unit problem (5), or more narrowly the issue of geographic scale, segregation levels vary between counties and metropolitan areas partly because of

their different sizes (6, 7). Thus, scale effect obscures inter-city, inter-regional, or inter-metropolitan comparisons. More importantly, a global approach fails to recognize the important variations of segregation levels across local areas or neighborhoods [e.g., Ref. (8–10)]. Unlike other areal units, the US census tracts are designed to be relatively homogeneous in regard to population characteristics, economic status, and living conditions (11). They are also delineated historically in accordance with uniform standards (12). Using them as local units (i.e., for the local approach) to examine the relationships between residential segregation and health is appropriate as they can portray people's daily environment more accurately than using other areal unit entities.

Equally important to the choice of areal units (i.e., the scale of analysis), the use of effective and meaningful segregation measures is critical in segregation studies. More than two decades ago, Massey and Denton (13) conducted an extensive and in-depth literature review on various segregation measures and classified them into five dimensions: (i) evenness (the differential distribution of population groups), (ii) exposure or, its counterpart, isolation (the potential interaction of population groups), (iii) concentration, similar to the concept of density (the distributional intensity of population groups), (iv) centralization (the dispersion of population groups with respect to the city center), and (v) clustering (the degree of spatial separation or proximity of population groups).



Because segregation measures without considering the spatial relationships are not effective [e.g., Ref. (14–16)], however, Reardon and O'Sullivan (17) concluded that evenness and isolation are the two distinct conceptual dimensions of segregation (Figure 1). Here, the centralization and concentration dimensions were omitted from the conceptual framework since both are regarded as specific subcategories of the evenness dimension (17). Johnston et al. (18) also concluded that the five conceptual dimensions of segregation should be reduced to two (i.e., evenness and isolation); they were unable to replicate Massey and Denton's empirical work in identifying the original five dimensions using the 1980–2000 US censuses. The dimensions of evenness and isolation are generally regarded to be distinctive, and therefore, segregation measures reflecting these two dimensions are preferable. By incorporating the concept from spatial statistics (19, 20), Wong (9, 10) developed a family of local spatial segregation indexes to capture the evenness and isolation dimensions of segregation at the neighborhood level.

Despite the conceptual importance of the local approach and thus the use of local spatial segregation indexes, they have rarely been incorporated into health research. A disconnect between the concept and measurement of segregation and its applications to health research would undermine the study of the role of residential segregation in health. In reference to the recent review article (4), for example, most previous studies based on a local approach have used the percentage of population belonging to a group (e.g., percent white and black) as a measure of segregation; among them,

a large body of literature has focused on adult mortality and pregnancy outcomes (e.g., low birth weight and/or preterm birth). Possibly, due to the adoptions of ineffective and insufficient segregation measures, the associations of residential segregation with adult mortality (21–26) and pregnancy outcomes (27–32) have been mixed. Drawing from a rather inconclusive body of literature (including the 12 studies listed above) nevertheless, Kramer and Hogue (4) suggested that the clustering (i.e., unevenness) and isolation dimensions of segregation may have a protective and adverse effect on health, respectively. While White and Borrell (33) conceived percentages as a “proxy” measure of segregation, percentages are not a measure of segregation because they cannot quantify how different population groups are distributed across space. Of particular importance in segregation studies, percentages cannot reflect the two distinct dimensions of segregation (13, 17, 18). From a critical point of view, the continued use of percentages as a “proxy” measure of segregation in health research would obscure the understanding of pathways by which residential segregation may shape health.

As Kramer and Hogue (4) argued, the evenness and isolation dimensions of segregation may have different effects on health; in other words, plausible protective and adverse (or null) effects of segregation may vary by health behaviors (e.g., diet, physical activity, and smoking) and health outcomes (e.g., cancer, obesity, and mortality), as well as by age, gender, race/ethnicity, and other sociodemographic characteristics. However, without incorporating effective and meaningful segregation measures in describing the neighborhood characteristics into health research, only limited (if not biased) knowledge can be gained from future studies. Much effort is, therefore, needed to build upon the conceptual and methodological foundations of measuring segregation established by demographers, geographers, and sociologists. In order to promote informative research, this paper offers several important conceptual remarks about measuring segregation, related methodological challenges, and practical approaches to set the stage for using segregation measurements in future studies. Two Midwestern US cities, St. Louis, MO, USA and Chicago, IL, USA, are used as examples to illustrate these points.

MATERIALS AND METHODS

DATA

Population characteristics by race/ethnicity, poverty status, and employment status at the census tract level were obtained from the 2005 to 2009 American Community Survey (ACS) for St. Louis, MO, USA (St. Charles County, St. Louis County, and St. Louis City) and Chicago, IL, USA (Cook County). Census tract data were used partly because both poverty and employment statuses were not available from the 2005 to 2009 ACS at the block group level. In this study, unpopulated census tracts were removed from the analysis. The 2000 US census tract boundary file was obtained from the US Census Bureau. Since census tract boundaries extend into rivers and/or include large ponds and lakes, such water bodies were also removed from the boundary file when the total land area (square kilometer) was recalculated in GIS (ArcGIS 10; ESRI Inc., Redlands, CA, USA). The population and geographic characteristics of these two Midwestern US cities are summarized in Table 1.

Table 1 | Description of the two Midwestern US cities.

		St. Louis, MO, USA	Chicago, IL, USA
Total land area ^a	(km ²)	2,918	2,433
Number of census tract ^b		340	1,327
Total population ^b		1,692,563	5,257,001
Non-Hispanic White ^b	(%)	70.0	45.2
Non-Hispanic Black ^b	(%)	23.3	25.3
Hispanic ^b	(%)	2.3	22.5
Other racial/ethnic groups ^b	(%)	4.4	7.0
Below poverty ^b	(%)	11.2	14.9
Unemployed ^b	(%)	7.2	9.3

^aDerived from the GIS recalculation (not including bodies of water) by the authors.

^bDerived from the 2005 to 2009 American Community Survey (ACS).

LOCAL SPATIAL SEGREGATION MEASURES

Traditional segregation measures are aspatial, meaning that when the populations of any two areal units are swapped, the segregation level of the entire region reflected by these aspatial measures remain unchanged. That is, these measures fail to consider population characteristics in the surrounding neighborhoods. Local segregation measures, which provide a segregation value for each local unit within a region, share the same deficiency that it is independent of the population characteristics in neighboring units. By incorporating the concept of modeling local spatial autocorrelation (19, 20), Wong (15) suggested using the function $c_{ij}(\cdot)$ to modify aspatial segregation indexes into spatial segregation indexes. In spatial autocorrelation studies, $c_{ij}(\cdot)$ is the element of a binary matrix where “1” indicates areal units i and j are neighbors, and “0” otherwise. Different from the $c_{ij}(\cdot)$ typically used in spatial autocorrelation studies, i can equal to j and thus $c_{ii} = 1$. Therefore, integrating the function $c_{ij}(\cdot)$ with some segregation indexes provides a means to include populations in neighboring units that can account for the potential spatial interaction of population groups across unit boundaries. This is the concept of composite population (15). For example, the composite population count of group G in areal unit i (cg_i) is modeled as

$$cg_i = \sum_j c_{ij}g_j$$

where g_j is the population count of group G in areal unit j . In other words, a composite population count refers to the population count in areal unit i plus the population counts in its neighboring units j .

The general concept of composite population count is to include the population of neighboring units in evaluating the population of a reference unit. The specific version of c_{ij} in the initial implementation of composite population counts was a binary function of neighbors such that the population of a neighboring unit is included or not. Feitosa et al. (8) used a kernel function to derive the local population intensity, the same concept as the composite population count such that populations farther away are weighted less, while closer are weighted more. To a large degree, this kernel estimator captures the distance decay effect commonly

considered in geographical research, reflecting the general fact that interaction declines with increasing distances. However, this formulation is slightly different from the “population density of the local environment” suggested by Reardon and O’Sullivan (17). Their function used a spatial density kernel to enumerate the population count of an areal unit, but these “(w)ighted density values depend on the spatial arrangement (distance between cells) and on the areas of the spatial units (cells),” and therefore, it is more affected by the sizes of areal units when spatially aggregated data are used (8).

While these subsequent developments employed more elegant spatial weighting schemes, but are more difficult to implement than the binary weights in the composite population counts, the importance is the notion that enumeration unit boundaries do not delimit the spatial interaction of population groups, and neighboring populations need to be considered in measuring the levels of segregation. Based upon the concept of composite population counts, Wong (9, 10) introduced a family of local spatial segregation indexes: the local spatial dissimilarity index (SD_i), the local spatial entropy-based diversity index (SH_i), and the local spatial isolation index (S_i). Massey and Denton (13) claimed that the traditional dissimilarity index (D) is best to capture the evenness dimension, while Reardon and Firebaugh (34) advocated the use of diversity index (H). H is another measure for the evenness dimension, which is also commonly referred to as the Shannon index in biology and ecology. Lieberman’s isolation index (P^*) has been regarded as the standard for the isolation dimension (35). By definition, Wong’s indexes are the local spatial versions of these global aspatial indexes, corresponding to the two distinct dimensions of segregation (i.e., evenness and isolation). These specifications are given as follows:

White–black dissimilarity (SD_{i^*wb})

$$SD_{i^*wb} = \left| \frac{cw_i}{CW} - \frac{cb_i}{CB} \right| \quad (1)$$

where cw_i and cb_i are the composite population counts of whites and blacks in areal unit i , respectively, and CW and CB are the composite population counts of whites and blacks for the entire study area, respectively. This is the local spatial version of the popular D .

White–others dissimilarity (SD_{i^*wo})

$$SD_{i^*wo} = \left| \frac{cw_i}{CW} - \left(\frac{ct_i - cw_i}{CT - CW} \right) \right| \quad (2)$$

Black–others dissimilarity (SD_{i^*bo})

$$SD_{i^*bo} = \left| \frac{cb_i}{CB} - \left(\frac{ct_i - cb_i}{CT - CB} \right) \right| \quad (3)$$

where ct_i is the composite population count of total population in areal unit i , and CT is the composite population count of total population for the entire study area. Whereas Eq. 1 compares the distributions of two groups, Eqs 2 and 3 compare one group with the remainder of the population. These are also the local spatial versions of the popular D .

Racial/ethnic diversity (SH_i)

$$SH_i = - \sum_k^n \left(\frac{cp_{ik}}{ct_i} \right) \ln \left(\frac{cp_{ik}}{ct_i} \right) \quad (4)$$

where cp_{ik} is the composite population count of mutually exclusive group k in areal unit i (e.g., whites, blacks, Hispanics, . . . n). This is the local spatial version of the popular H .

White isolation (S_{i^*wo})

$$S_{i^*wo} = 1 - \left(\frac{ct_i - cw_i}{T - W} \right) \quad (5)$$

Black isolation (S_{i^*bo})

$$S_{i^*bo} = 1 - \left(\frac{ct_i - cb_i}{T - B} \right) \quad (6)$$

where W , B , and T are the population counts of whites, blacks, and total population for the entire study area, respectively. Here, the original formulations (9) were simplified in this study by replacing the neighborhood operators with the composite population counts. These are the modified local spatial versions of the popular P^* .

Note that either equations 5 or 6 can be applied to model the spatial interaction of below and above the poverty status as poverty isolation, as well as of unemployed and employed (among the civilian non-institutional population aged 16 years and older) status as unemployment isolation.

These local spatial segregation measures were computed in R (36). The local spatial dissimilarity measures (derived from SD_i) and the local spatial entropy-based diversity measures (derived from SH_i) were standardized by their maximum values, while the local spatial isolation measures (derived from S_i) were standardized by their range, such that all measures are bounded between

0 and 1. These standardizations are justifiable as the purpose of this study is to examine intra-urban variations, not inter-city differences.

ANALYSIS

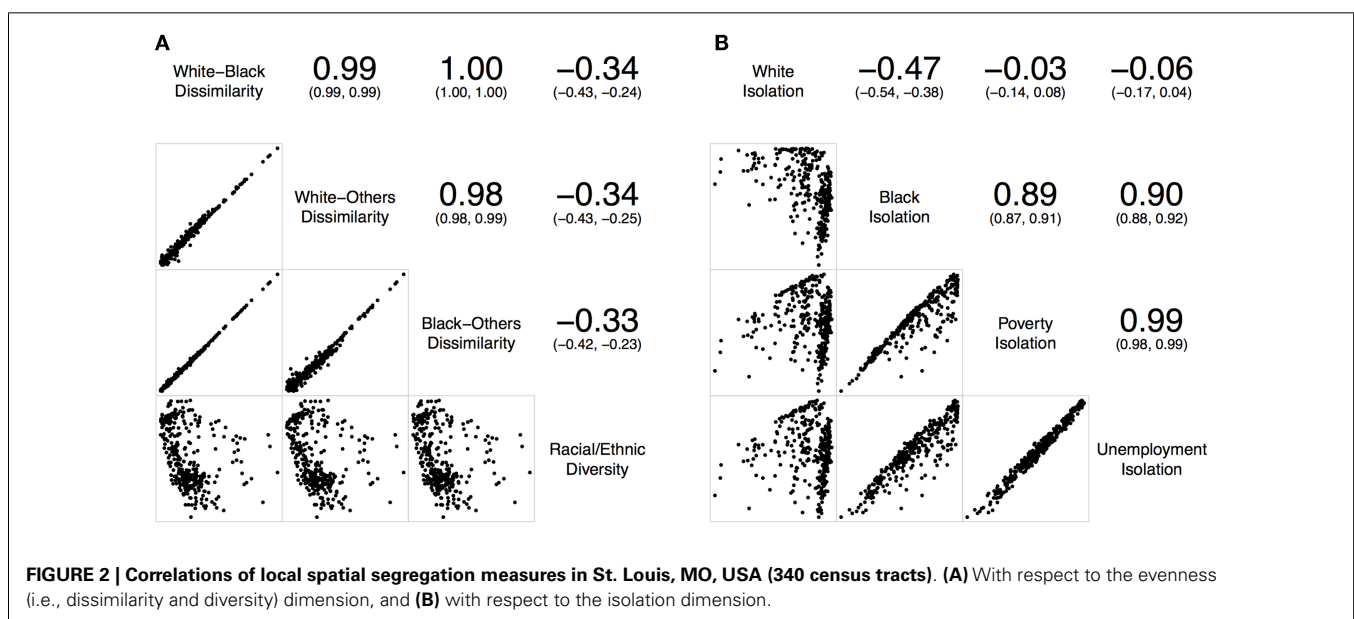
To examine the relationships of local spatial segregation measures derived from above, a series of correlation statistics (37) were conducted in R (38) for St. Louis, MO, USA (Figure 2) and Chicago, IL, USA (Figure 3). Correlations and scatterplot matrixes were used to display the relationships. The upper off-diagonal panels show the correlation coefficients with associated 95% confidence intervals (in parentheses), and the lower off-diagonal panels show the scatter plots. The geographical distributions of local spatial segregation measures with respect to the evenness (Figures 4 and 5) and isolation (Figures 6 and 7) dimensions within the two Midwestern US cities are shown in maps. A quantile classification scheme was used to display the levels of local segregation.

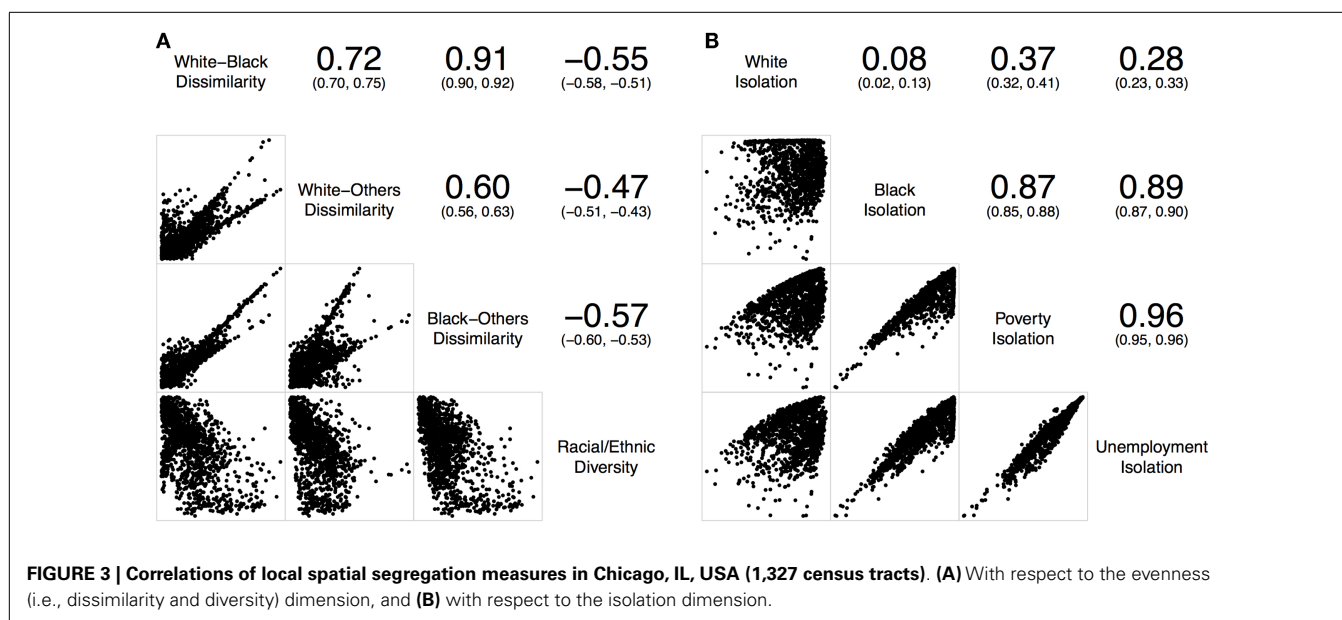
For demonstration purposes, the geographical distributions of percent white, black, Hispanic, and other racial/ethnic groups are also shown for St. Louis, MO, USA (Figure 8) and Chicago, IL, USA (Figure 9).

RESULTS

Table 1 shows the geographic and population characteristics of the study area. In brief, St. Louis, MO, USA and Chicago, IL, USA are in the same geographic region with a similar total land area. However, these two Midwestern US cities were different in terms of their population sizes, number of census tracts, and racial/ethnic compositions. In addition, the poverty and unemployment rates were slightly higher in Chicago, IL, USA than those in St. Louis, MO, USA.

Figure 2A shows the correlations of different local spatial segregation measures with respect to the evenness (i.e., dissimilarity and diversity) dimension in St. Louis, MO, USA. The white–black, white–others, and black–others dissimilarity measures were highly and positively correlated with one another ($0.98 \leq r \leq 1.00$).





However, these dissimilarity measures were not correlated with the racial/ethnic diversity measure ($-0.33 \leq r \leq -0.34$). On the other hand, **Figure 2B** shows the correlations of different local spatial segregation measures with respect to the isolation dimension. While the white isolation measure was moderately, but negatively correlated with the black isolation measure ($r = -0.47$), it was not correlated with either the poverty ($r = -0.03$) or unemployment ($r = -0.06$) isolation measures. However, the black, poverty, and unemployment isolation measures were highly and positively correlated with one another ($0.89 \leq r \leq 0.99$).

Similarly, **Figure 3A** shows the correlations of different local spatial segregation measures with respect to the evenness (i.e., dissimilarity and diversity) dimension in Chicago, IL, USA. The white-black dissimilarity measure was moderately and positively correlated with the white-others ($r = 0.72$) and highly and positively correlated with the black-others ($r = 0.91$) dissimilarity measures where the white-others dissimilarity measure was moderately correlated with the black-others dissimilarity measure ($r = 0.60$). In addition, these dissimilarity measures were moderately and negatively correlated with the racial/ethnic diversity measure ($-0.47 \leq r \leq -0.57$). On the other hand, **Figure 3B** shows the correlations of different local spatial segregation measures with respect to the isolation dimension. The white isolation measure was not correlated with the black ($r = 0.08$), poverty ($r = 0.37$), or unemployment ($r = 0.28$) isolation measures. However, the black, poverty, and unemployment isolation measures were highly and positively correlated with one another ($0.87 \leq r \leq 0.96$).

Taken together, **Figures 2** and **3** consistently show that: (i) the two-group-based dissimilarity measures do not capture the local variation of racial/ethnic segregation as the multiple-group-based diversity measure does, (ii) white isolation was neither equal to nor the inverse of black isolation, and (iii) black isolation occurred conjointly with the poverty and unemployment isolation. However, it is important to note that **Figures 2** and **3** also show slightly inconsistent degrees of correlation. In general, St. Louis, MO,

USA has higher correlations among the dissimilarity measures than those in Chicago, IL, USA, but their correlations with the racial/ethnic diversity measure were lower in St. Louis, MO, USA than those in Chicago, IL, USA (**Figure 2A** versus **Figure 3A**). In addition, St. Louis, MO, USA has higher, but modest, correlations between the white and black isolation measures than those in Chicago, IL, USA. However, the correlations among the white, poverty, and unemployment isolation measures were higher in Chicago, IL, USA than those in St. Louis, MO, USA (**Figure 2B** versus **Figure 3B**). These inconsistent degrees of correlation between the two Midwestern US cities were primarily due to the difference in racial/ethnic compositions (**Table 1**); the larger proportions of Hispanic and other racial/ethnic groups in Chicago, IL, USA contributes more to the spatial patterns than those in St. Louis, MO, USA.

Results of correlation analysis in **Figures 2A** and **3A** are also manifested spatially. **Figures 4** and **5** show the geographic distributions of local spatial evenness (i.e., dissimilarity and diversity) measures in St. Louis, MO, USA and Chicago, IL, USA, respectively. In **Figures 4A–C**, the dissimilarity measures between white-black, white-others, and black-others exhibit very similar spatial patterns in St. Louis, MO, USA, with clusters of high segregation in the city center (downtown) and patches in the western outskirts. The high degrees of concurrences between these dissimilarity measures were primarily driven by the spatial pattern of blacks as they constitute the dominant minority group in the region. However, such high degrees of resemblance between these dissimilarity measures were not found in Chicago, IL, USA (**Figures 5A–C**). Differences between these two Midwestern US cities are visually recognizable, as besides whites and blacks, Hispanics population constitutes a major proportion in Chicago, IL, USA (**Table 1**). As pointed out above in the correlation analysis, the dissimilarity and diversity measures do not have an inverse relation (**Figures 2A** and **3A**). By comparing **Figures 4D** and **5D** with **Figures 4A–C** and **5A–C**, respectively, they confirm that the racial/ethnic diversity measure

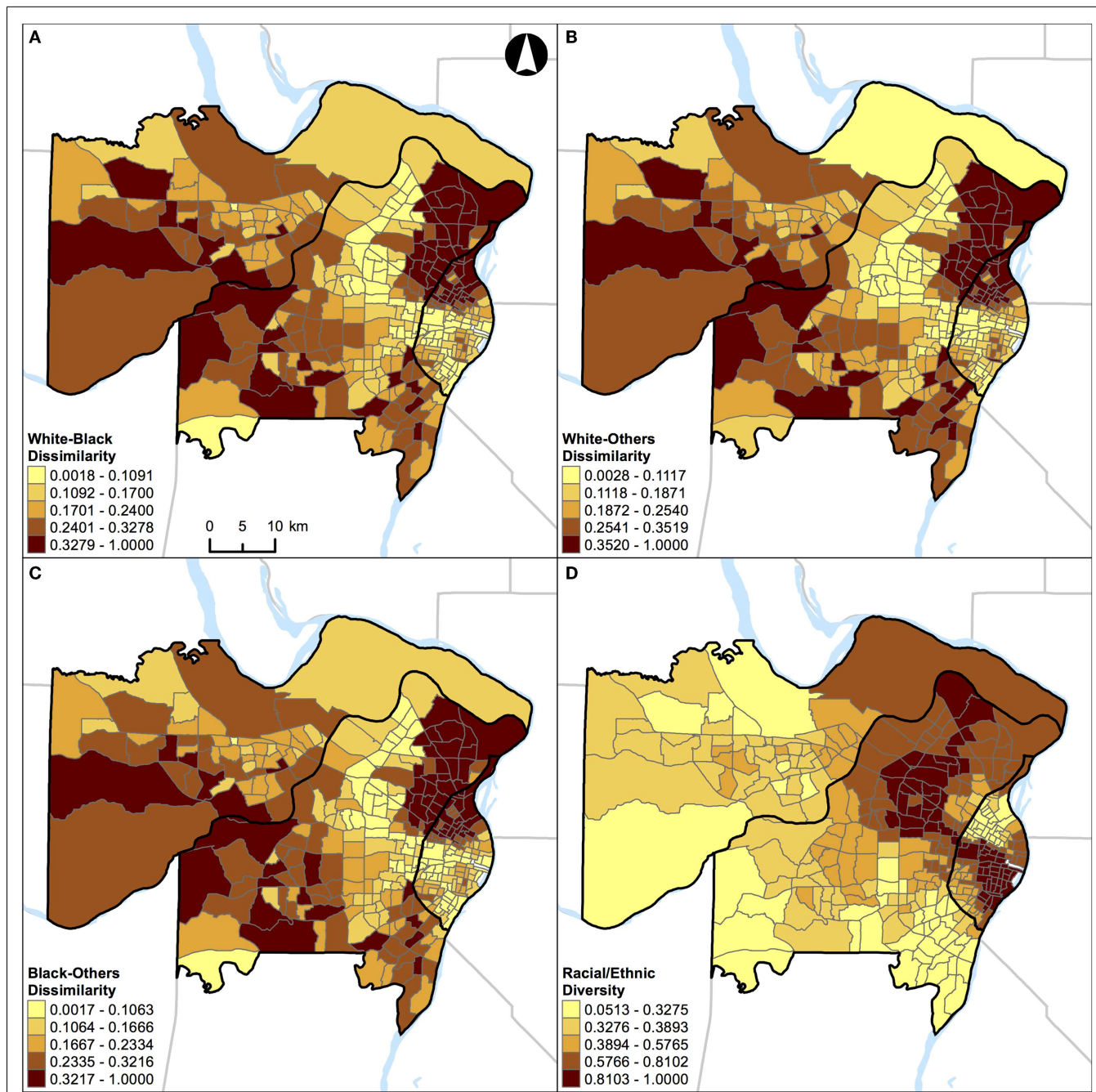
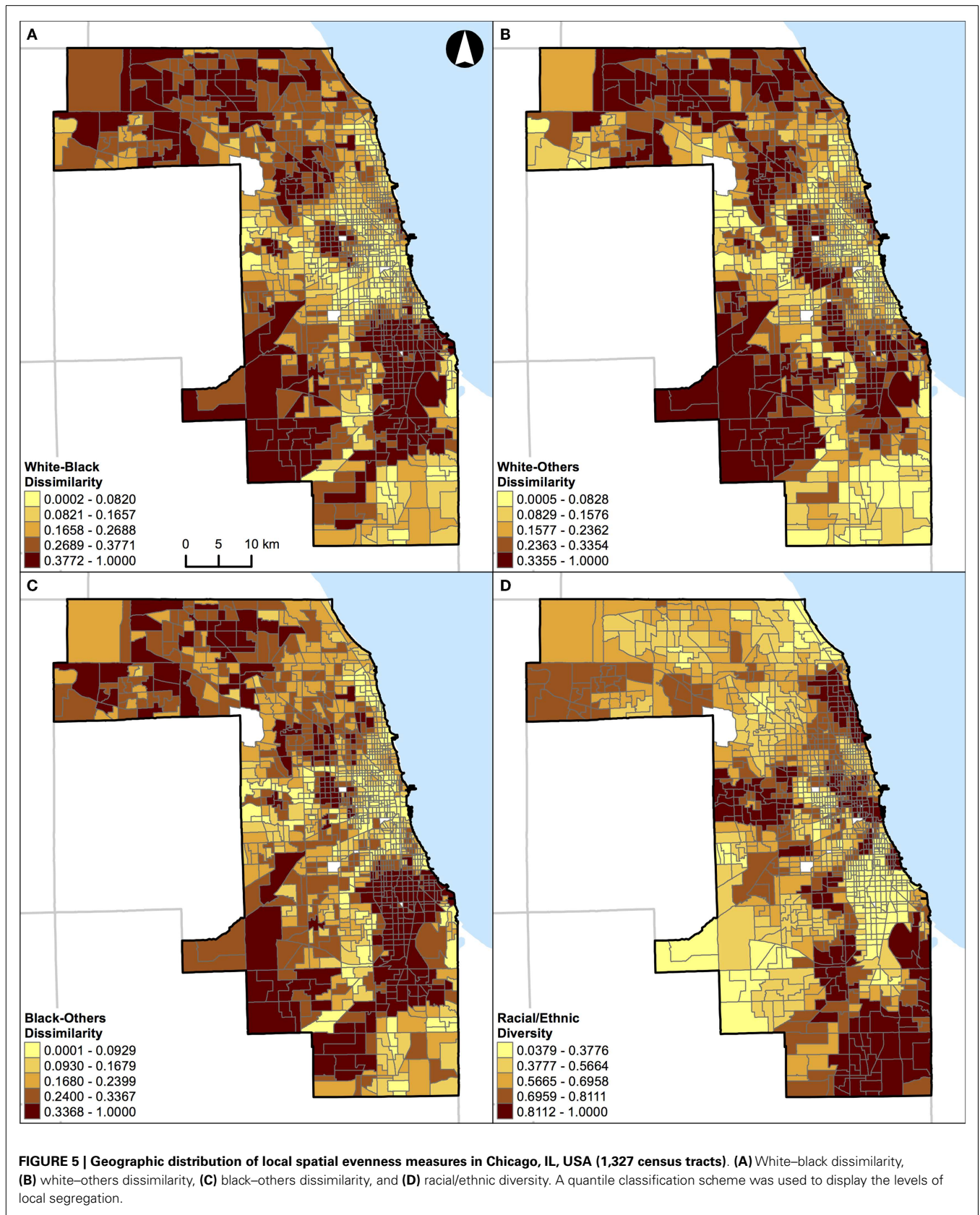


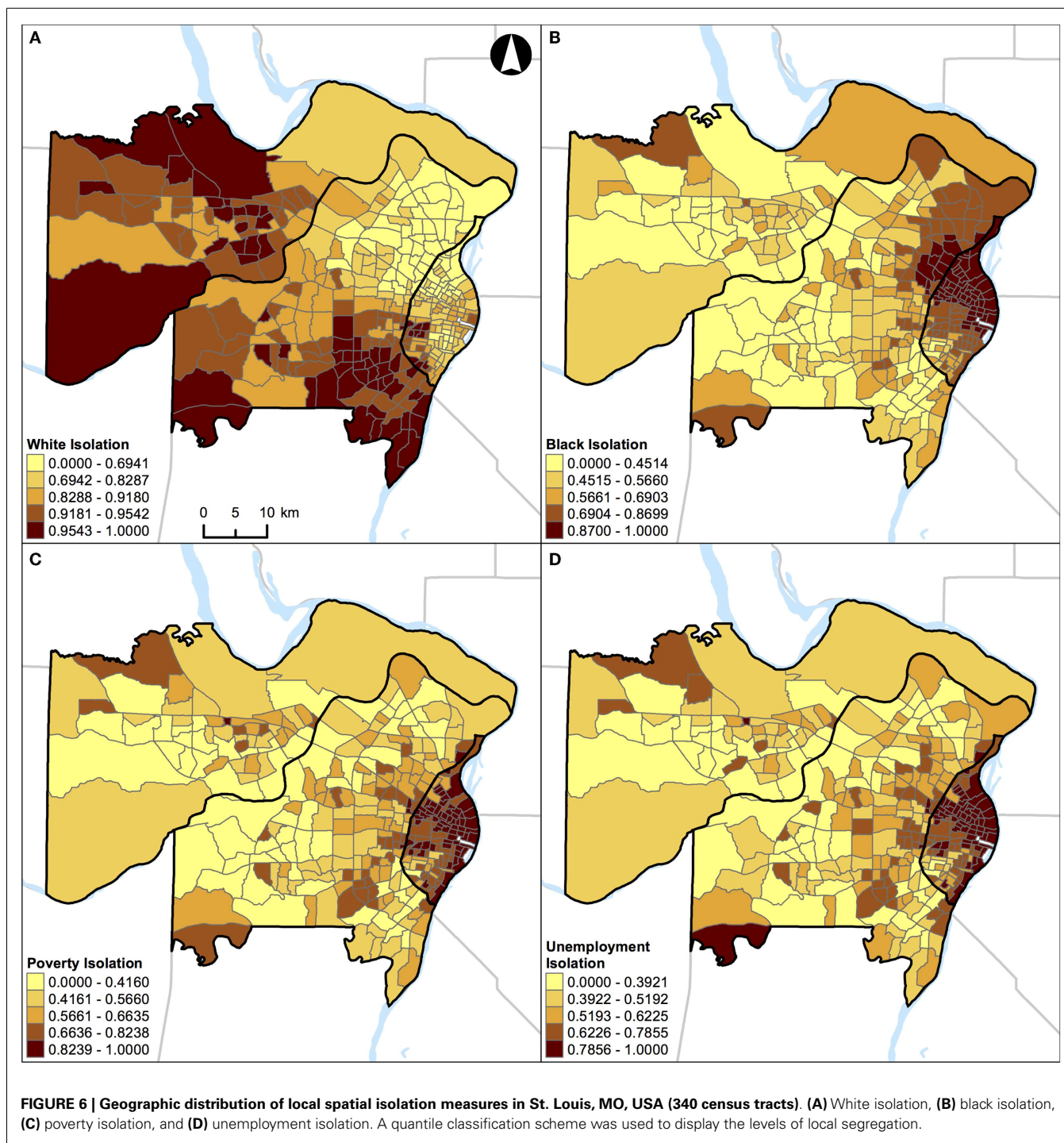
FIGURE 4 | Geographic distribution of local spatial evenness measures in St. Louis, MO, USA (340 census tracts). (A) White–black dissimilarity, **(B)** white–others dissimilarity, **(C)** black–others dissimilarity, and **(D)** racial/ethnic diversity. A quantile classification scheme was used to display the levels of local segregation.

captures more than just the opposite of dissimilarity; areas with the highest (or lowest) values in **Figures 4A–C** and **5A–C** do not always have the lowest (or highest) values in **Figures 4D** and **5D**, respectively.

In addition, results of correlation analysis in **Figures 2B** and **3B** are also manifested spatially. **Figures 6** and **7** show the geographic distributions of local spatial isolation measures in St. Louis, MO,

USA and Chicago, IL, USA, respectively. **Figures 6B–D** and **7B–D** exhibit very similar spatial patterns thereby reflecting the fact that blacks are the socioeconomically disadvantaged group in these two Midwestern US cities. Although the spatial patterns of white isolation (**Figures 6A** and **7A**) are different from those of the black, poverty, and unemployment isolation (**Figures 6B–D** and **7B–D**, respectively), they are not reversed images of each other. In fact,





some areas, such as the southwestern and northwestern corners of St. Louis, MO, USA seem to have relatively high isolation levels in all situations (Figures 6A–D). On the other hand, in the northeastern corner of Chicago, IL, USA, it has both relatively high levels of white and unemployment isolation (Figures 7A,D). In order to explain these spatial patterns and their relationships, further research is needed to understand in detail the interaction between racial/ethnic and socioeconomic segregations. However,

it is beyond the scope of this study. When data about health are brought into the analysis, Figure 4 through Figure 7 can serve as the foundations for exploratory spatial data analysis (ESDA) to assist the formulations of hypotheses (39, 40) in examining the relationships between residential segregation and health.

Unlike the segregation measures described above, percentage of population belonging to a group (e.g., percent black) cannot be used as a “proxy” measure of segregation (33) in future

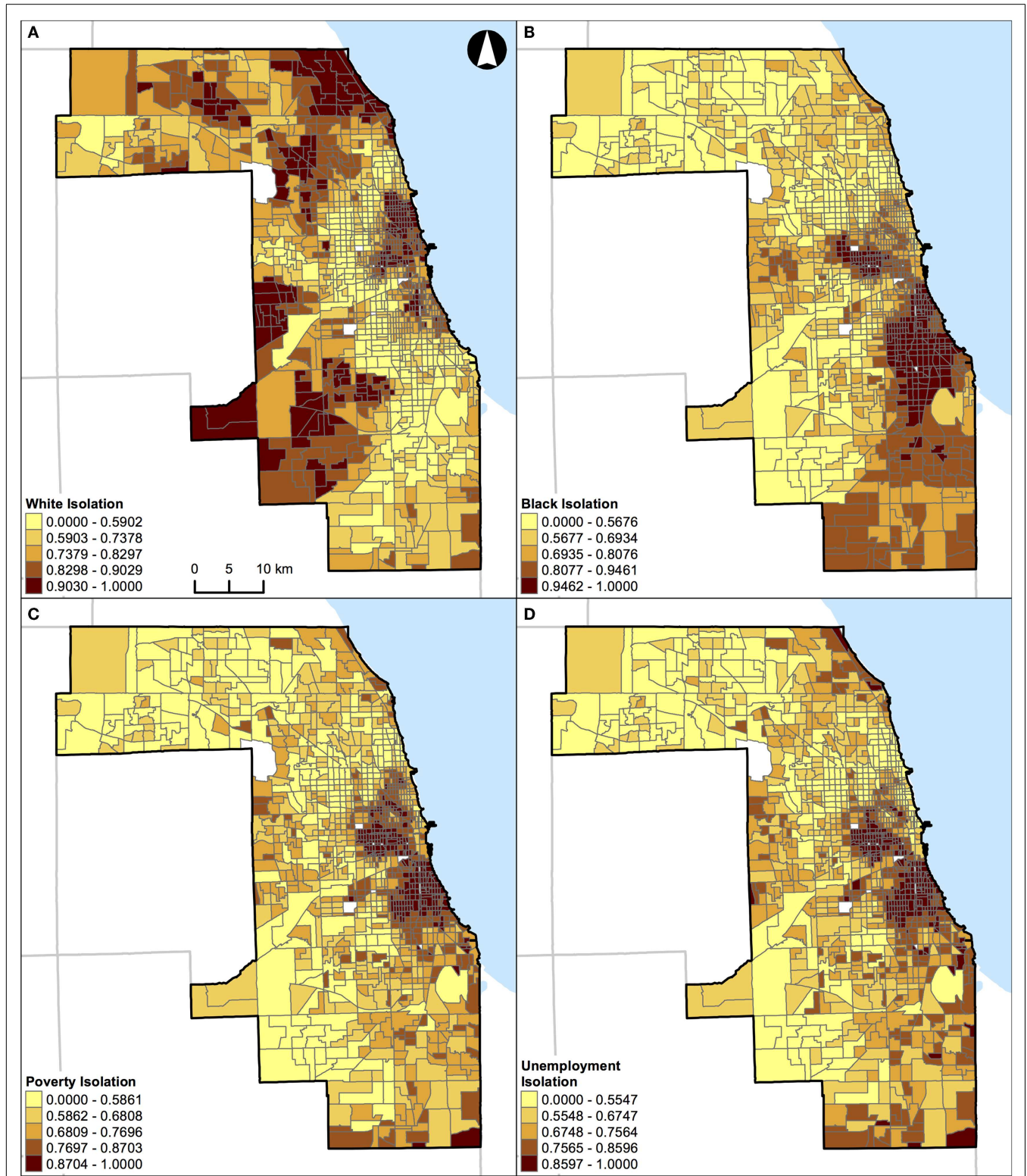
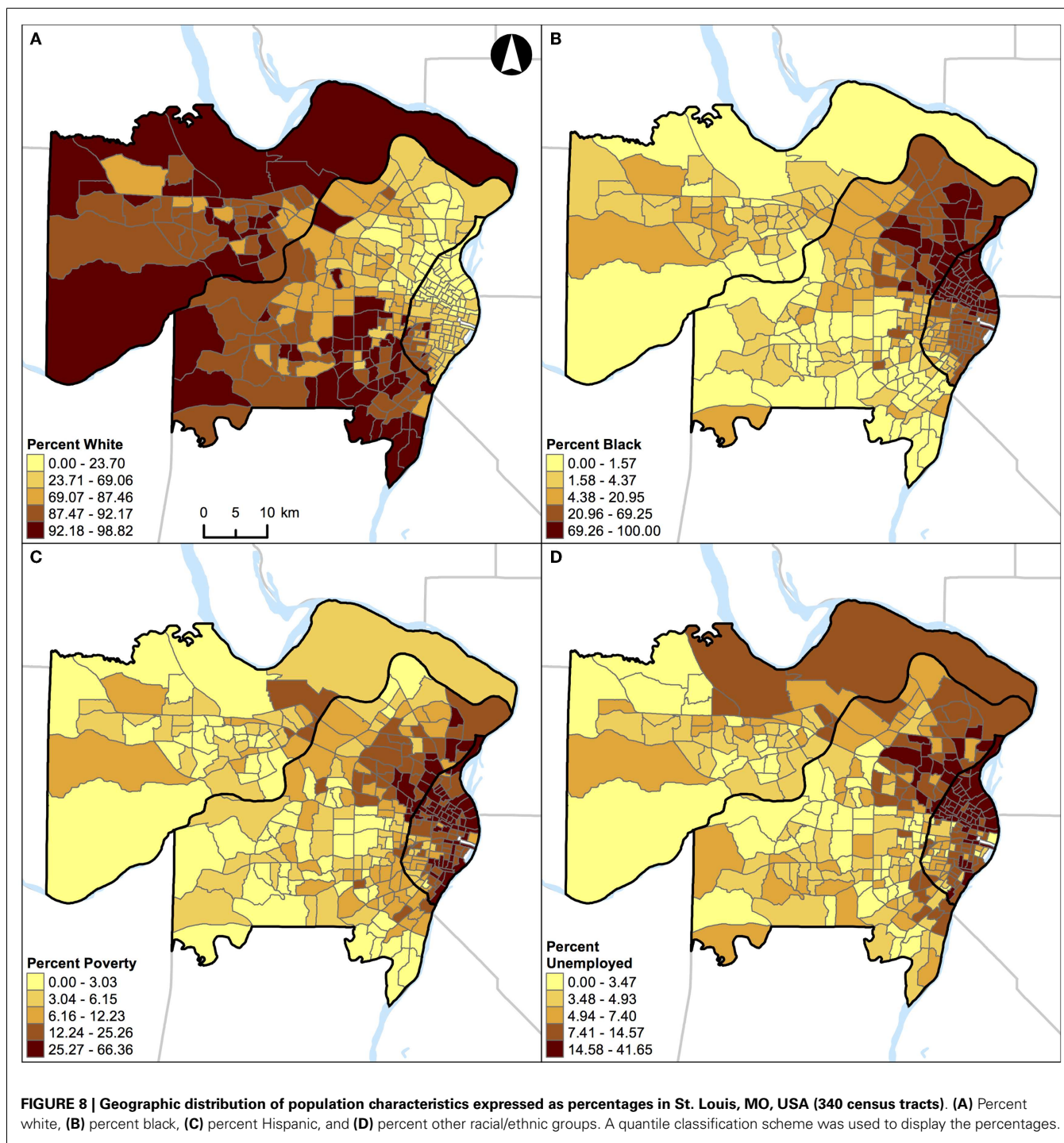


FIGURE 7 | Geographic distribution of local spatial isolation measures in Chicago, IL, USA (1,327 census tracts). (A) White isolation, **(B)** black isolation, **(C)** poverty isolation, and **(D)** unemployment isolation. A quantile classification scheme was used to display the levels of local segregation.



research. This argument can be illustrated by examining the geographic distributions of percent white, black, Hispanic, and other racial/ethnic groups in the two Midwestern US cities (Figures 8 and 9). For example, in St. Louis, MO, USA, a lower percentage of white (Figure 8A) corresponds to higher percentages of other population groups in different parts of the region: black in the northeastern parts (Figure 8B), Hispanic in the north central and southeastern parts (Figure 8C), and other racial/ethnic

groups in the central parts (Figure 8D). In addition, higher percentages of black (Figure 8B), Hispanic (Figure 8C), and other racial/ethnic groups (Figure 8D) co-occur in the north central and southeastern parts of St. Louis, MO, USA. Similarly, in Chicago, IL, USA, a lower percentage of white (Figure 9A) corresponds to higher percentages of other population groups in different parts of the region: black in the central and southeastern parts (Figure 9B), Hispanic in the northern, central, and southeastern

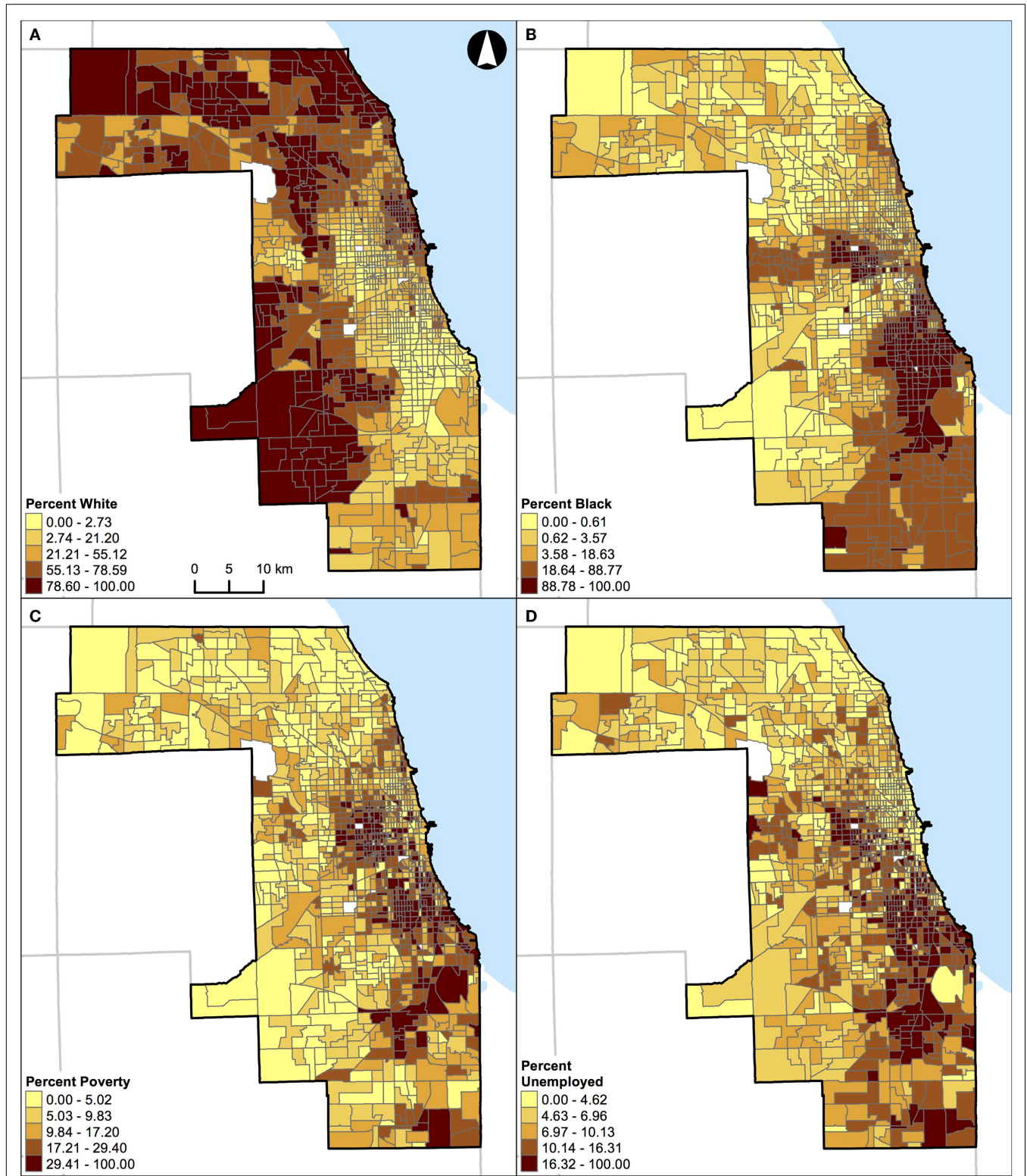


FIGURE 9 | Geographic distribution of population characteristics expressed as percentages in Chicago, IL, USA (1,327 census tracts). (A) Percent white, (B) percent black, (C) percent Hispanic, and (D) percent other racial/ethnic groups. A quantile classification scheme was used to display the percentages.

parts (**Figure 9C**), and other racial/ethnic groups along the shore of Lake Michigan and in the northern parts (**Figure 9D**). In addition, higher percentages of white (**Figure 9A**), black (**Figure 9B**), Hispanic (**Figure 9C**), and other racial/ethnic groups (**Figure 9D**) co-occur along the shore of Lake Michigan as well as in the central and southern parts of Chicago, IL, USA. If percentages were used to measure the level of segregation, such usages would suggest that: (i) a lower percentage of white would refer not only to a higher percentage of black (i.e., the inverse thereof) but also to higher percentages of Hispanic and other racial/ethnic groups, but these high percentages may not be true, and (ii) a higher percentage of a racial/ethnic group (particularly among the minority groups) would refer to both a racially/ethnically dominated and integrated neighborhood. Taken together, **Figures 8** and **9** corroborate the fact that percentages cannot reflect the two distinct dimensions of segregation (13, 17, 18).

One of the central principles in segregation studies is that measures need to quantify how different population groups are distributed across areal units. Putting aside the aspatial nature for a moment, the segregation indexes have been developed by demographers, geographers, and sociologists to measure the extent to which two or more groups are distributed across areal units within a given region. Simple percentages cannot capture the between-unit relationship as implicitly accounted for in most segregation measures, aspatial or spatial indexes alike. For these reasons, using local spatial segregation indexes to reflect the evenness and isolation dimensions are preferable in future studies.

DISCUSSION

In this study, St. Louis, MO, USA and Chicago, IL, USA were examined. They are in the same geographical region with similar total land areas, but different population characteristics (**Table 1**). As pointed out above, the correlations of local spatial segregation measures showed consistent patterns in these two Midwestern US cities (**Figures 2** and **3**), and these correlations were also manifested spatially (**Figure 4** through **Figure 7**). As a practical guide to segregation studies, **Figures 8** and **9** demonstrated that percentages cannot be used as a measure of segregation. These results, in turn, highlight two important remarks about the use of local spatial segregation indexes in health research.

First, the use of local spatial entropy-based diversity index (SH_i) is recommended for measuring the local variation of racial/ethnic segregation with respect to the evenness dimension. Both the dissimilarity and diversity indexes have been classified as measures of evenness among the dimensions of segregation (13, 17). From a conceptual point of view, these two measures are the inverse of each other. As shown in **Figures 2A**, **3A**, **4**, and **5**, however, such an expectation does not generally hold. By and large, the dissimilarity index (D) has become one of the most popular (if not considered to be the standard) measure of segregation. Nevertheless, the use of D in segregation studies has long been criticized for its limitations [e.g., Ref. (41–43)] since it was first introduced nearly six decades ago by Duncan and Duncan (44). In fact, Cortese et al. (45) demonstrated some of the systematic biases in D more than three decades ago. Among existing measures of segregation, the entropy-based diversity index (H) has been determined as a superior measure of the evenness dimension (34, 46). Because H

is aspatial in nature, the use of its spatial version (SH) is logical in segregation studies (9, 10, 17). In a world of increasing globalization, the US has become markedly more racially and ethnically diverse (mainly, due to the increases in Hispanic and Asian populations) during the recent decades (47). Unlike the past, the situations of multiple racial/ethnic groups are the norms rather than the exception (15, 34). Future research on the relationship between residential segregation and health, therefore, should consider using SH_i to measure the evenness dimension of racial/ethnic segregation in order to account for the spatial dimension and the multiracial and ethnic nature of the contemporary US society.

Second, the use of local spatial isolation index (S_i) is recommended not only for measuring the local variation of racial/ethnic segregation but also for socioeconomic (e.g., poverty and unemployment) segregation with respect to the isolation dimension. From a structural view of racial phenomena in the US, segregation of blacks has been considered as the institutional manifestations of racism: the racial organization of the society is structured by the underlying psychological, cultural, social, economic, and political phenomena over time. In other words, the observed racial disparity is induced and maintained because a society possesses a racialized social system that determines the relationship and interaction between races (48). Particularly for blacks, limited educational and employment opportunities, redlining and housing discrimination, and adverse psychosocial distress are some of the societal factors coupled with residential segregation (2, 3). Due to the multifaceted nature of racial phenomena, segregation of blacks may ultimately constrain their wealth accumulation and upward social mobility. At the national level, poor blacks have been segregated from other racial/ethnic and income groups, and the magnitude of segregation had weakened only slightly between 1970 and 2000 (49). Overall, a major shift of blacks into less-segregated (or integrated) areas has not occurred (50). Taken together, these historic and societal factors are likely to play a role in constraining poor and unemployed blacks to live in certain neighborhoods, but not for whites (**Figures 2B**, **3B**, **6**, and **7**). Future research on the relationship between residential segregation and health, therefore, should consider using S_i to measure the isolation dimension of racial/ethnic and socioeconomic segregations to account for the spatial dimension and the multifaceted nature of population distribution in the US.

Despite the number of studies conducted to date (4), the wide adoptions of ineffective and insufficient segregation measures reflect the ongoing trend that the methodological advancements in segregation studies achieved by demographers, geographers, and sociologists have not been adequately translated into health research. Namely, the percentage of population belonging to a group (e.g., percent black) has been widely used as a measure of segregation in most previous studies based on a local approach. To our knowledge, there has not been a study that used the local spatial segregation indexes for measuring racial/ethnic diversity and socioeconomic (e.g., poverty and unemployment) isolation in health research; only four studies used S_i for measuring black isolation in a form of continuous and log-transformed variables (27, 51) or dichotomous categorical variables (28, 52). As a supplemental note, two studies incorporated the segregation indexes based

on spatial kernels (17): black–others dissimilarity and black isolation as continuous and binary variables (53), and black isolation as a categorical variable (54). However, no study has incorporated the segregation indexes developed by Feitosa et al. (8). Notwithstanding the methodological differences (as noted earlier), the segregation indexes developed by Reardon and O’Sullivan (17) and Feitosa et al. (8) are comparable to, but methodologically more elegant than those developed by Wong (9, 10).

Toward the use of SH_i and S_i as covariates in health research, however, there are technical and theoretical challenges that need to be explored in future studies, especially in the context of linear regression analysis. That is, these two local spatial segregation measures do not necessarily follow a normal distribution and may be skewed or highly skewed. In some instances, neither applying the traditional transformations (e.g., square root, log, and inverse) nor the Box–Cox (i.e., parametric power) transformation (55) can achieve normality. Even when normality has been achieved, data transformations can introduce difficulties in interpreting the results as they alter the nature of the variable. As a means to handle skewed or highly skewed distributions, continuous variables have commonly been categorized into quantiles (most often tertiles or quartiles) based on various approaches [e.g., Ref. (56–61)]. Dichotomizing continuous variables has been widely regarded as an inappropriate practice that generates rather than solves problems (62, 63). However, it is equally important to recognize that categorization of continuous variables has also been criticized for its unrealistic assumption of homogeneity within categories, and its rather arbitrary data-driven cut points used to define categories that may create difficulties to compare results across studies (64, 65). To date, there is no theoretical or empirical basis for determining the optimal cut points in categorizing the local spatial segregation measures.

Further research may be needed to determine thresholds for which the local spatial segregation measures have protective or adverse effects on health. However, it may not be feasible to identify the threshold levels that apply to different localities, particularly for the evenness dimension. For example, a growing racial/ethnic diversity related to the increases in Hispanic and Asian populations varies from region to region in the US (47). Also, within-city and within-suburban sorting of racial/ethnic populations does not occur evenly or uniformly across the US (66). Even between St. Louis, MO, USA and Chicago, IL, USA, which are roughly 400 km apart, their racial/ethnic compositions are quite different (Table 1); in particular, the proportion of Hispanics in St. Louis, MO, USA is much lower than that in Chicago, IL, USA. As a result, the increase of racial/ethnic diversity (derived from SH_i) would mostly reflect white–black (i.e., racially) integrated neighborhoods in St. Louis, MO, USA, whereas such an increase would reflect white–black–Hispanic (i.e., racially and ethnically) integrated neighborhoods in Chicago, IL, USA. This, in turn, suggests that the potential interaction of racial/ethnic groups and thus their experience in social environments are not alike; the same level of racial/ethnic diversity measure across cities does not necessarily mean that the same degree and nature of racial/ethnic diversity are present. For these reasons, the threshold effects of local spatial segregation on health should be explored with careful consideration and clear justifications. In health research, generalizability

and transportability are two important components that should be kept in mind.

Another challenge of using SH_i and S_i in a regression framework is related to the processes of how they are computed. Regardless of whether they are derived using the simplistic composite population concept (9, 10) or the elegant spatial kernels (8, 17), the basic principle of local spatial segregation indexes is to remove the enumeration boundaries as the absolute barriers to inter-group interaction by aggregating populations across adjacent or contiguous neighborhoods. As noted earlier, this is a more realistic portrayal of the social interaction among neighbors in our daily lives than that of such interaction to occur only within the confined unit boundary. The aggregation processes smooth the distribution of population spatially by which the smoothed data lead to a better detection of the systematic trends in a given area. However, such aggregation processes can also magnify the level of spatial autocorrelation (i.e., residuals that vary systematically over space), which violates the independence assumption. For example, LeSage (67) demonstrated how ignoring the spatial configuration of geographic data in a linear regression model can produce unstable parameter estimates and yield unreliable statistical significance testing results. Although a certain level of spatial autocorrelation is expected, particularly when the data are from an urban context, both SH_i and S_i are likely to accentuate the magnitude and structure of positive spatial autocorrelation. Different types of spatial regression models have been introduced to account for spatial dependencies, such as the spatial error or spatial lag models and the eigenvector-based spatial filtering method [e.g., Ref. (68–70)]. Their implementations, nevertheless, create another layer of complexity in statistical modeling. To our knowledge, no study examined the relationships between residential segregation and health using the local spatial segregation measures (derived from SH_i and S_i) in spatial regression models.

These technical and theoretical challenges can restrict the use of SH_i and S_i in a regression model when an outcome of interest is a continuous variable. Without due consideration, the distribution of these two local spatial segregation measures may violate the assumptions of linearity, normality, homoscedasticity, and/or independence. Nonetheless, an outcome of interest is often, or can be classified into, a binary variable in health research (e.g., obesity and smoking status). Unlike a linear regression model, a logistic regression model (as well as other types of generalized linear models) does not require normality in the distributions of covariates and homoscedasticity (i.e., homogeneity of variance), and has less stringent requirements (71). For this reason, as long as the linearity in the logit assumption is satisfied, logistic regression that accounts for spatial autocorrelation can be used to examine the relationships between the local spatial segregation measures (derived from SH_i and S_i) and health-related outcomes. Various spatial logistic models have been developed in recent years, but the results and conclusions may vary between different models [e.g., Ref. (72)]. As a preliminary step for fitting regression models to spatial data, a model selection should be based on the comparative analysis of different models (73).

More appropriately, from a statistical perspective, multilevel logistic models should be used instead of spatial logistic models. Here, multilevel (i.e., hierarchical, mixed, nested, mixed-effect,

or random-effect) models refer to regression models that combine traditionally distinct individual and ecological models, and to overcome the limitations in focusing only on one level (74, 75). The use of multilevel models is a necessity in exploring the relationships between residential segregation and health (1) because both individualistic and ecological fallacies can lead to inaccurate (if not misleading) conclusions (76). In order to conduct an informative analysis, covariates at multiple levels should be taken into consideration in explaining health-related outcomes. At least four studies have examined the relationship between residential segregation and health by using the spatial isolation index in multilevel models (27, 28, 51, 54). Hence, the use of multilevel models is recommended as the fundamental statistical approach in future studies.

By taking advantage of modern computational and statistical methods, a more sophisticated approach to commonly used multilevel models is the use of (Bayesian) generalized additive mixed models [e.g., Ref. (77)]. Given the complexity of such statistical models (77), however, they must be considered with care.

LIMITATIONS

Despite the potential importance of using SH_i and S_i in health research as illustrated above, these two local spatial segregation indexes share common concerns with other spatial analytical techniques and segregation studies. Among them, two inherent limitations warrant mentioning. For one, both SH_i and S_i are subject to the boundary or edge effect, which is typically unavoidable when using areal and geographic data. Such an effect introduces bias into the identification of spatial distribution and the parameter estimates of spatial process (78). That is, in almost any given geographical and health studies, a boundary has to be demarcated as the study area: a city, a metropolitan area, or a region defined by one or more counties in the US. While only subunits within the study area are the concerns in the analysis, units outside of the demarcated boundary would be ignored completely. In other words, areal units within the study areas are clipped from the rest of the geography to form an island, isolated from the rest in the analytical setting. The exception is in a unique setting where the study area is entirely surrounded by a large body of water (e.g., studies conducted within the Hawaii islands). Since areal association, geographic distribution, and spatial interaction extend beyond the demarcated boundary, when the study area is “lifted” from its surrounding geography in the analysis, the measures or statistics computed would be biased. In fact, all units within the entire study area would be affected, but the effects are stronger for areal units closer to the border than those closer to the center of the study area. Inevitably, the local spatial segregation measures (SH_i and S_i) would be affected by the boundary or edge effect.

Several solutions have been proposed to address the boundary or edge effect during the past decades. Nonetheless, all of them have misgivings, and they cannot be implemented easily [e.g., Ref. (79)], or can they fully solve the problem. One rather simple approach, which was not implemented in this study, is to include a buffer zone around the study area. For instance, additional areas to the west of the two Midwestern US cities can be used to create the buffer zone. However, the Mississippi River, which runs along the east side of St. Louis, MO, USA, physically separates the

city from East St. Louis, IL, USA, another city to the east. While a couple of bridges and public transportation systems connect the Missouri and Illinois sides, limited transit options to cross the river substantially reduce the spatial interaction of population groups between the two sides of the river. In the computation of local spatial segregation measures using SH_i and S_i , all areal units are treated as spatially continuous regardless of such boundary lines being the street or river. In reality, however, the spatial interaction of population groups and their experience in the social environments are not the same for people crossing the street versus the river. Given the physical separation between the two sides of the river, it may not be appropriate to extend a buffer zone beyond the Mississippi River. Otherwise, a careful consideration is needed for creating a buffer zone into the Illinois side. Unlike St. Louis, MO, USA, Lake Michigan lies on the east side of Chicago, IL, USA, and thus, creating a buffer zone will not be appropriate and feasible.

In general, the larger the buffer zone, the less will be the boundary or edge effect. The ideal size of a buffer zone is to ensure that all areal units within the study area are not affected. However, such a size is very difficult to determine; it will be partly dependent upon the spatial analytical technique adopted and the data used in a study, which may or may not be place specific. Because a gold standard does not exist, a very careful consideration is needed instead in handling areal and geographic data (80, 81). One simple guideline is to identify the nature of the spatial relationship to be involved, and then to determine the size of the buffer zone in minimizing the boundary or edge effect to an acceptable level. In the context of SH_i and S_i , the function $c_{ij}(\cdot)$ adopted to implement the concept of composite population involves only the immediate neighboring units (15). For practical purposes, therefore, a buffer zone including the first-order adjacent units along the study area will be sufficient for using these two local spatial segregation indexes to measure the levels of segregation.

Lastly, a local approach (and in fact to a large degree, a global approach) to measure socioeconomic (e.g., poverty and unemployment) segregation would encounter a challenge in terms of data quality. In particular, this issue pertains to the US situation. After 2000, US Census no longer gathers detailed information on household's socioeconomic status and living conditions through the so-called long form in previous decennial censuses. Replacing the long form is the ACS, which is a continuous measurement program surveying US population, and is the major source of socioeconomic and housing data. Due to the nature and design of ACS, however, its estimates for smaller geographical units (e.g., census tracts and block groups) may not be reliable (with relatively large margin of errors) as compared with data from past decennial censuses. Nevertheless, how the quality of ACS estimates may affect the computation of segregation measures in general, and SH_i and S_i in specific have not been investigated. As is to be expected, a curtain of uncertainty will be casted over the socioeconomic segregation measures when the ACS data are used in future studies.

CONCLUSION

Segregation is the extent to which individuals of various groups occupy and experience different social environments. As the condition involves more than one group, measuring the levels of

segregation needs to account for the spatial interaction of different population groups. Otherwise, segregation measures without accounting for spatial relationships would leave out the essence of segregation. Unlike the global segregation measures that overlook the important variations at the local level, local segregation measures draw attentions to the situations at the neighborhood scale. In particular, two local spatial segregation indexes highlighted in this paper provide effective and meaningful measurements of the two distinct dimensions of segregation: (i) the local spatial entropy-based diversity index (SH_i) for the evenness dimension, and (ii) the local spatial isolation index (S_i) for the isolation dimension. From an analytical point of view, the use of SH_i will help elucidate the relationship between racial/ethnic integration (or, its counterpart, racial/ethnic similarity) and health, whereas the use of S_i will help elucidate the relationships between racial/ethnic and/or socioeconomic isolation and health. These two local spatial segregation indexes can be used in ESDA, assisting the formulations of hypotheses to be tested, and examining the relationship between residential segregation and health. However, they have rarely been incorporated into health research. Hence, future studies should explore the use of SH_i and S_i to better understand both the protective and adverse effects of residential segregation on health.

AUTHOR CONTRIBUTIONS

Masayoshi Oka conceptualized the study, obtained the data, carried out the analyses, and drafted the manuscript. David W. S. Wong contributed to the conceptualization of the study, drafted some subsections, and revised the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

This work was supported by the National Cancer Institute (NCI) Centers for Transdisciplinary Research on Energetics and Cancer (TREC) (U54CA155496) and by the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP/2007–2013)/ERC Starting Grant HeartHealthy-Hoods Agreement No. 336893. The NCI and ERC had no further role in the study design, collection, analysis, and interpretation of data, writing of the article, or decision for publication. The opinions and assertions contained herein are the private ones of the authors and are not considered as official or reflecting the views of the National Institute of Health and the European Union.

REFERENCES

- Acevedo-Garcia D, Lochner KA, Osypuk TL, Subramanian SV. Future directions in residential segregation and health research: a multilevel approach. *Am J Public Health* (2003) **93**(2):215–21. doi:10.2105/AJPH.93.2.215
- Williams DR. Race, socioeconomic status, and health: the added effects of racism and discrimination. *Ann N Y Acad Sci* (1999) **896**:173–88. doi:10.1111/j.1749-6632.1999.tb08114.x
- Williams DR, Collins C. Racial residential segregation: a fundamental cause of racial disparities in health. *Public Health Rep* (2001) **116**(5):404–16. doi:10.1016/S0033-3549(04)50068-7
- Kramer MR, Hogue CR. Is segregation bad for your health? *Epidemiol Rev* (2009) **31**(1):178–94. doi:10.1093/epirev/mxp001
- Openshaw S. *The Modifiable Areal Unit Problem*. Norwich: Geo Books (1984).
- Krupka DJ. Are big cities more segregated? Neighbourhood scale and the measurement of segregation. *Urban Stud* (2007) **44**(1):187–97. doi:10.1080/00420980601023828
- Reardon SF, Matthews SA, O'Sullivan D, Lee BA, Firebaugh G, Farrell CR, et al. The geographic scale of metropolitan racial segregation. *Demography* (2008) **45**(3):489–514. doi:10.1353/dem.0.0019
- Feitosa FF, Câmara G, Monteiro AM, Koschitzki T, Silva MP. Global and local spatial indices of urban segregation. *Int J Geogr Inf Sci* (2007) **21**(3):299–323. doi:10.1080/13658810600911903
- Wong DW. Modeling local segregation: a spatial interaction approach. *Geogr Environ Model* (2002) **6**(1):81–97. doi:10.1080/13615930220127305
- Wong DW. A local multi-dimensional approach to evaluate changes in segregation. *Urban Geogr* (2008) **29**(5):455–72. doi:10.2747/0272-3638.29.5.455
- United States Census Bureau. *Geographic Areas – Definitions*. Washington, DC: American Community Survey Office (2013).
- Krieger N. A century of census tracts: health & the body politic (1906–2006). *J Urban Health* (2006) **83**(3):355–61. doi:10.1007/s11524-006-9040-y
- Massey DS, Denton NA. The dimensions of residential segregation. *Soc Forces* (1988) **67**(2):281–315. doi:10.1093/sf/67.2.281
- Morrill RL. On the measure of geographic segregation. *Geogr Res Forum* (1991) **11**:25–36.
- Wong DW. Measuring multiethnic spatial segregation. *Urban Geogr* (1998) **19**(1):77–87. doi:10.2747/0272-3638.19.1.77
- Wong DW. Formulating a general spatial segregation measure. *Prof Geogr* (2005) **57**(2):285–94. doi:10.1111/j.0033-0124.2005.00478.x
- Reardon SF, O'Sullivan D. Measures of spatial segregation. *Sociol Methodol* (2004) **34**(1):121–62. doi:10.1111/j.0081-1750.2004.00150.x
- Johnston R, Poulsen M, Forrest J. Ethnic and racial segregation in U.S. Metropolitan areas, 1980–2000: the dimensions of segregation revisited. *Urban Aff Rev* (2007) **42**(4):479–504. doi:10.1177/1078087406292701
- Anselin L. Local indicators of spatial association – LISA. *Geogr Anal* (1995) **27**(2):93–115. doi:10.1111/j.1538-4632.1995.tb00338.x
- Getis A, Ord JK. The analysis of spatial association by use of distance statistics. *Geogr Anal* (1992) **24**(3):189–206. doi:10.1111/j.1538-4632.1992.tb00261.x
- Collins CA. Racism and health: segregation and causes of death amenable to medical intervention in major U.S. Cities. *Ann N Y Acad Sci* (1999) **896**:396–8. doi:10.1111/j.1749-6632.1999.tb08152.x
- Fang J, Madhavan S, Bosworth W, Alderman MH. Residential segregation and mortality in New York City. *Soc Sci Med* (1998) **47**(4):469–76. doi:10.1016/S0277-9536(98)00128-2
- Guest AM, Almgren G, Hussey JM. The ecology of race and socioeconomic distress: infant and working-age mortality in Chicago. *Demography* (1998) **35**(1):23–34. doi:10.2307/3004024
- Inagami S, Borrell LN, Wong MD, Fang J, Shapiro MF, Asch SM. Residential segregation and Latino, black and white mortality in New York City. *J Urban Health* (2006) **83**(3):406–20. doi:10.1007/s11524-006-9035-8
- Jackson SA, Anderson RT, Johnson NJ, Sorlie PD. The relation of residential segregation to all-cause mortality: a study in black and white. *Am J Public Health* (2000) **90**(4):615–7. doi:10.2105/AJPH.90.4.615
- LeClere FB, Rogers RG, Peters K. Ethnicity and mortality in the United States: individual and community correlates. *Soc Forces* (1997) **76**(1):169–98. doi:10.2307/2580322
- Grady SC. Racial disparities in low birthweight and the contribution of residential segregation: a multilevel analysis. *Soc Sci Med* (2006) **63**(12):3013–29. doi:10.1016/j.socscimed.2006.08.017
- Grady SC, McLafferty SL. Segregation, nativity, and health: reproductive health inequalities for immigrant and native-born black women in New York City. *Urban Geogr* (2007) **28**(4):377–97. doi:10.2747/0272-3638.28.4.377
- Pickett KE, Collins JW, Masi CM, Wilkinson RG. The effects of racial density and income incongruity on pregnancy outcomes. *Soc Sci Med* (2005) **60**(10):2229–38. doi:10.1016/j.socscimed.2004.10.023
- Masi CM, Hawkey LC, Piotrowski ZH, Pickett KE. Neighborhood economic disadvantage, violent crime, group density, and pregnancy outcomes in a diverse, urban population. *Soc Sci Med* (2007) **65**(12):2440–57. doi:10.1016/j.socscimed.2007.07.014
- Suocoff CA, Upchurch DM. Neighborhood context and the risk of childbearing among metropolitan-area black adolescents. *Am Sociol Rev* (1998) **63**(4):571–85. doi:10.2307/2657268
- Vinikoor LC, Kaufman JS, MacLehose RE, Lارايا BA. Effects of racial density and income incongruity on pregnancy outcomes in less segregated communities. *Soc Sci Med* (2008) **66**(2):255–9. doi:10.1016/j.socscimed.2007.08.016

33. White K, Borrell LN. Racial/ethnic residential segregation: framing the context of health risk and health disparities. *Health Place* (2011) **17**(2):438–48. doi:10.1016/j.healthplace.2010.12.002
34. Reardon SF, Firebaugh G. Measures of multigroup segregation. *Sociol Methodol* (2002) **32**(1):33–67. doi:10.1111/1467-9531.00110
35. Lieberman S. An asymmetrical approach to segregation. In: Peach C, Robinson V, Smith S, editors. *Ethnic Segregation in Cities*. London: Croom-Helm (1981). p. 61–82.
36. Core Team R. *R: A Language and Environment for Statistical Computing* Version 3.1.0. Vienna: R Foundation for Statistical Computing (2014).
37. Friendly M. Corrgrams: exploratory display for correlation matrices. *Am Stat* (2002) **56**(4):316–24. doi:10.1198/000313002533
38. Wright K. *Corrgram: Plot a Correlogram*. R package version 1.4 (2012).
39. Anselin L. How (not) to lie with spatial statistics. *Am J Prev Med* (2006) **30**(2S):S3–6. doi:10.1016/j.amepre.2005.09.015
40. Unwin A, Unwin D. Exploratory spatial data analysis with local statistics. *J Roy Stat Soc D Sta* (1998) **47**(3):415–21. doi:10.1111/1467-9884.00143
41. Reiner TA. Racial segregation: a comment. *J Regional Sci* (1972) **12**(1):137–48. doi:10.1111/j.1467-9787.1972.tb00286.x
42. Winship C. The desirability of using the index of dissimilarity or any adjustment of it for measuring segregation: reply to falk, cortese, and cohen. *Soc Forces* (1978) **57**(2):717–20. doi:10.1093/sf/57.2.717
43. Zelder RE. Racial segregation: a reply. *J Regional Sci* (1972) **12**(1):149–53. doi:10.1111/j.1467-9787.1972.tb00286.x
44. Duncan OD, Duncan B. A methodological analysis of segregation indexes. *Am Sociol Rev* (1955) **20**(2):210–7. doi:10.2307/2088328
45. Cortese CF, Falk FR, Cohen JK. Further considerations on the methodological analysis of segregation indices. *Am Sociol Rev* (1976) **41**(4):630–7. doi:10.2307/2094840
46. White MJ. Segregation and diversity measures in population distribution. *Popul Index* (1986) **52**(2):198–221. doi:10.2307/3644339
47. Iceland J. Beyond black and white: metropolitan residential segregation in multi-ethnic America. *Soc Sci Res* (2004) **33**(2):248–71. doi:10.1016/S0049-089X(03)00056-5
48. Bonilla-Silva E. Rethinking racism: toward a structural interpretation. *Am Sociol Rev* (1997) **62**(3):465–80. doi:10.2307/2657316
49. Fischer MJ. The relative importance of income and race in determining residential outcomes in U.S. Urban areas, 1970–2000. *Urban Aff Rev* (2003) **38**(5):669–96. doi:10.1177/1078087403038005003
50. Logan JR, Stults BJ, Farley R. Segregation of minorities in the metropolis: two decades of change. *Demography* (2004) **41**(1):1–22. doi:10.1353/dem.2004.0007
51. Grady SC, Ramirez IJ. Mediating medical risk factors in the residential segregation and low birthweight relationship by race in New York City. *Health Place* (2008) **14**(4):661–7. doi:10.1016/j.healthplace.2007.10.011
52. White K, Borrell LN, Wong DW, Galea S, Ogedegbe G, Glymour MM. Racial/ethnic residential segregation and self-reported hypertension among US- and foreign-born blacks in New York City. *Am J Hypertens* (2011) **24**(8):904–10. doi:10.1038/ajh.2011.69
53. Kramer MR, Cooper HL, Drews-Botsch CD, Waller LA, Hogue CR. Do measures matter? Comparing surface-density-derived and census-tract-derived measures of racial residential segregation. *Int J Health Geogr* (2010) **9**:29. doi:10.1186/1476-072X-9-29
54. Greer S, Casper M, Kramer M, Schwartz G, Hallisey E, Holt J, et al. Racial residential segregation and stroke mortality in Atlanta. *Ethn Dis* (2011) **21**(4):437–43.
55. Box GE, Cox DR. An analysis of transformations. *J Roy Stat Soc B Met* (1964) **26**(2):211–52.
56. Cureton EE. The upper and lower twenty-seven per cent rule. *Psychometrika* (1957) **22**(3):293–6. doi:10.1007/BF02289130
57. D'Agostino RB. The 27 percent rule revisited. *Educ Psychol Meas* (1975) **35**(1):47–50. doi:10.1177/001316447503500105
58. Gelman A, Park DK. Splitting a predictor at the upper quarter or third and the lower quarter or third. *Am Stat* (2008) **62**(4):1–8. doi:10.1198/016214507000000000
59. O'Brien SM. Cutpoint selection for categorizing a continuous predictor. *Bio-metrics* (2004) **60**(2):504–9. doi:10.1111/j.0006-341X.2004.00196.x
60. Preacher KJ, Rucker DD, MacCallum RC, Nicewander WA. Use of the extreme groups approach: a critical reexamination and new recommendations. *Psychol Methods* (2005) **10**(2):178–92. doi:10.1037/1082-989X.10.2.178
61. Ross J, Weitzman RA. The twenty-seven per cent rule. *Ann Math Stat* (1964) **35**(1):214–21. doi:10.1214/aoms/1177703745
62. Cohen J. The cost of dichotomization. *Appl Psych Meas* (1983) **7**(3):249–53. doi:10.1177/014662168300700301
63. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med* (2006) **25**(1):127–41. doi:10.1002/sim.2331
64. Bennette C, Vickers A. Against quantiles: categorization of continuous variables in epidemiologic research, and its discontents. *BMC Med Res Methodol* (2012) **12**:21. doi:10.1186/1471-2288-12-21
65. Weinberg CR. How bad is categorization? *Epidemiology* (1995) **6**(4):345–7. doi:10.1097/00001648-199507000-00002
66. Fischer MJ. Shifting geographies: examining the role of suburbanization in blacks' declining segregation. *Urban Aff Rev* (2008) **43**(4):475–96. doi:10.1177/1078087407305499
67. LeSage JP. Regression analysis of spatial data. *J Reg Anal Pol* (1997) **27**(2):83–94.
68. Anselin L. *Spatial Econometrics: Methods and Models*. Boston, MA: Kluwer Academic Publisher (1988).
69. Elliott P, Wakefield JC, Best NG, Briggs DJ. *Spatial Epidemiology: Methods and Applications*. New York, NY: Oxford University Press (2001).
70. Griffith DA. *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding Through Theory and Scientific Visualization*. Berlin: Springer-Verlag (2003).
71. Nelder JA, Wedderburn WM. Generalized linear models. *J Roy Stat Soc A Sta* (1972) **135**(3):370–84. doi:10.2307/2344614
72. Dormann CF, McPherson JM, Araújo MB, Bivand R, Bolliger J, Carl G, et al. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* (2007) **30**(5):609–28. doi:10.1111/j.2007.0906-7590.05171.x
73. Beale CM, Lennon JJ, Yearsley JM, Brewer MJ, Elston DA. Regression analysis of spatial data. *Ecol Lett* (2010) **13**(2):246–64. doi:10.1111/j.1461-0248.2009.01422.x
74. Diez Roux AV. Multilevel analysis in public health research. *Annu Rev Public Health* (2000) **21**(1):171–92. doi:10.1146/annurev.publhealth.21.1.171
75. Diez Roux AV. A glossary for multilevel analysis. *J Epidemiol Community Health* (2002) **56**(8):588–94. doi:10.1136/jech.56.8.588
76. Subramanian SV, Jones K, Kaddour A, Krieger N. Revisiting Robinson: the perils of individualistic and ecologic fallacy. *Int J Epidemiol* (2009) **38**(2):342–60. doi:10.1093/ije/dyn359
77. Wood SN. *Generalized Additive Models: An Introduction with R*. Boca Raton, FL: Chapman and Hall/CRC Press (2006).
78. Griffith DA. The boundary value problem in spatial statistical analysis. *J Regional Sci* (1983) **23**(3):377–87. doi:10.1111/j.1467-9787.1983.tb00996.x
79. Griffith DA, Amrhein CG. An evaluation of correction techniques for boundary effects in spatial statistical analysis: traditional methods. *Geogr Anal* (1983) **15**(4):352–60. doi:10.1111/j.1538-4632.1983.tb00794.x
80. Griffith DA. Towards a theory of spatial statistics. *Geogr Anal* (1980) **12**(4):325–39. doi:10.1111/j.1538-4632.1980.tb00040.x
81. Griffith DA. Toward a theory of spatial statistics: another step forward. *Geogr Anal* (1987) **19**(1):69–82. doi:10.1111/j.1538-4632.1987.tb00114.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 28 May 2014; accepted: 30 July 2014; published online: 25 August 2014.

Citation: Oka M and Wong DWS (2014) Capturing the two dimensions of residential segregation at the neighborhood level for health research. *Front. Public Health* 2:118. doi: 10.3389/fpubh.2014.00118

This article was submitted to *Epidemiology*, a section of the journal *Frontiers in Public Health*.

Copyright © 2014 Oka and Wong. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.