



## Comparing proxy and formal measures of county-level racial isolation in race-stratified models: A case study in Tennessee, 2005–2014

Masayoshi Oka<sup>a,\*</sup>, Faustine Williams<sup>b</sup>, Martin Whiteside<sup>c</sup>

<sup>a</sup> Department of Management, Faculty of Management, Josai University, Sakado, Saitama, Japan

<sup>b</sup> Division of Intramural Research, National Institute on Minority Health and Health Disparities, National Institutes of Health, Bethesda, MD, USA

<sup>c</sup> Tennessee Cancer Registry, Tennessee Department of Health, Nashville, TN, USA

### ARTICLE INFO

#### Keywords:

Racial isolation  
Residential segregation  
Breast cancer condition  
Multilevel analysis

### ABSTRACT

The purpose of this study was to understand whether proxy measures of county-level racial isolation (based on racial compositions) would yield similar results as the formal measures of county-level racial isolation (derived from the isolation index of  $P^*$ ). White (non-Hispanic White) and Black (non-Hispanic Black or African American) women residing in the State of Tennessee, USA, and diagnosed with a non-invasive or invasive breast cancer were considered as the study population. Individual-level variables were obtained from the Tennessee Cancer Registry data for the period between 2005 and 2014 (46,983 White women and 7,967 Black women), and county-level variables were obtained from the American Community Survey data for the periods of 2005–2009 and 2010–2014 (95 counties). Using breast cancer condition (non-invasive versus invasive) as the binary outcome of interest, a series of multilevel logistic regression analyses was conducted separately by race. After controlling for individual-level socio-demographic characteristics, proxy measure of county-level White isolation and county-level median household income were not associated with breast cancer condition, but formal measure of county-level White isolation was associated with lower odds of having an invasive breast cancer among White women. On the other hand, neither proxy nor formal measure of county-level Black isolation was associated with breast cancer condition, but county-level median household income was associated with lower odds of having an invasive breast cancer among Black women. These results suggest that using a proxy and formal measure of racial isolation may yield different results, and race-stratified analyses would be helpful for understanding a differential effect of racial isolation on Whites and Blacks. While more detailed examinations are needed in future studies, possible explanations on and reasons behind these findings are discussed.

### 1. Introduction

A growing number of research has been conducted in the United States (US) to examine a potential role of residential segregation in cancer (e.g., breast cancer, colorectal cancer, lung cancer, and prostate cancer) since the mid-2000s (Landrine et al., 2017). The underlying motivation for this trend is based on the premise that the pervasive and persistent patterns of health disparities between Blacks (non-Hispanic Black alone or African Americans) and Whites (non-Hispanic White alone) may largely be a function of residential segregation (e.g., Landrine & Corral, 2009; Williams & Collins, 2001). This notion stems from observed patterns in which most Blacks live, work, and age under unfavorable social and physical environments compared to their White

counterparts irrespective of the socioeconomic status (e.g., Charles, 2003; Massey & Fischer, 2000; Williams, 1999). Viewing the pervasiveness and persistence of most Blacks residing (in essence, “trapped”) in areas of concentrated poverty, Williams and Collins (2001) conceived residential segregation as a fundamental cause of racial disparities in health.

After reviewing 17 studies (primarily on breast cancer) published between 2006 and 2016, Landrine et al. (2017) concluded that Blacks residing in segregated areas may lead to later-stage diagnosis of breast and lung cancers, higher mortality rates and lower survival rates from breast and lung cancers, and higher cumulative cancer risks associated with exposure to ambient air toxics. These adverse associations of residential segregation with cancer-related outcomes (Landrine et al., 2017)

*Abbreviations:* White, non-Hispanic White alone; Black, non-Hispanic Black alone or African American.

\* Corresponding author. Department of Management, Faculty of Management, Josai University, 1-1 Keyakidai, Sakado City, Saitama Prefecture, 350-0295, Japan.

E-mail address: [moka@alumni.harvard.edu](mailto:moka@alumni.harvard.edu) (M. Oka).

<https://doi.org/10.1016/j.ssmph.2022.101146>

Received 18 February 2022; Received in revised form 8 June 2022; Accepted 9 June 2022

Available online 14 June 2022

2352-8273/© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

mirror the results of those with various health-related outcomes (e.g., all-cause mortality, cause-specific mortality, birth weight, preterm birth, body mass index, smoking and self-rated health) reviewed by [Kramer and Hogue \(2009\)](#) and [K. White and Borrell \(2011\)](#). While these findings are based on cross-sectional studies ([Kramer & Hogue, 2009](#); [White & Borrell, 2011](#); [Landrine et al., 2017](#)), relatively consistent associations collectively corroborate the essence of [Williams and Collins' \(2001\)](#) arguments that residential segregation plays a role in creating conditions inimical to health among most Blacks. However, one of the major limitations commonly found in those reviewed studies ([Kramer & Hogue, 2009](#); [White & Borrell, 2011](#); [Landrine et al., 2017](#)) is the different ways by which residential segregation has been measured. Most noticeably, proxy measures based on racial/ethnic compositions (i.e., proportions or percentages of racial/ethnic groups) and formal measures derived from segregation indexes have been used interchangeably to refer to the degrees of residential segregation.

[Oka and Wong \(2014\)](#) argued that a measure of residential segregation used in health research must rely on the most appropriate index for capturing the specific dimension. In short, [Massey and his colleagues \(1988; 1996\)](#) analyzed 20 indexes of residential segregation and classified them into five dimensions: evenness, isolation (or exposure), clustering, concentration (which is similar to the concept of density), and centralization. However, [Reardon and O'Sullivan \(2004\)](#) and [Johnston, Poulsen, and Forrest \(2007\)](#) later concluded that these five dimensions can be reduced to two dimensions: evenness-clustering and isolation-exposure. Note that [Reardon and O'Sullivan \(2004\)](#) classified the concentration and centralization dimensions as subcategories of the evenness dimension, and [Johnston et al. \(2007\)](#) were unable to replicate [Massey and his colleagues' earlier results](#). Among the indexes of residential segregation, the information theory index ( $H$ ), which was introduced by [Shannon \(1948a, b\)](#) or [Theil \(1972\)](#) depending on the fields of study, has been considered most appropriate for capturing the evenness dimension ([M. J. White, 1986](#); [Reardon & Firebaugh, 2002](#); [Reardon & O'Sullivan, 2004](#)). Widely regarded as the standard index, the isolation index of  $P^*$ , which was introduced by [Bell \(1954\)](#) and popularized by [Lieberman \(1981\)](#), has been considered most appropriate for capturing the isolation dimension ([Lieberman, 1981](#); [Morrill, 1991](#); [Reardon & O'Sullivan, 2004](#)). In reviewed studies ([Kramer & Hogue, 2009](#); [White & Borrell, 2011](#); [Landrine et al., 2017](#)), most common measures of residential segregation were intended to capture the isolation dimension.

Because proxy measures (i.e., proportions or percentages of racial/ethnic groups) do not capture any of the distinct dimensions of residential segregation identified by [Massey and his colleagues \(1988; 1996\)](#), using proxy and formal measures interchangeably could obscure scientific knowledge of the relationships between residential segregation and health ([Oka & Wong, 2014](#)). While [Oka and Wong \(2014\)](#) provided valuable conceptual and theoretical discussions, they did not empirically demonstrate the difference between these two types of measures. Therefore, the purpose of this study was to understand whether proxy and formal measures of racial isolation would yield different results. Using breast cancer condition (non-invasive versus invasive) as the binary outcome of interest, the differences between proxy and formal measures of county-level racial isolation in the State of Tennessee were examined by a series of multilevel logistic regression analyses to inquire into the appropriateness of [Oka and Wong's \(2014\)](#) claims.

## 2. Methods

In this study, individual-level and county-level measures were combined to portray a hierarchical structure of residential segregation and health, and then a series of multilevel analyses was conducted to explore whether proxy and formal measures of racial isolation would yield different results. A series of multilevel analyses was conducted separately by race to set forth a differential effect of racial isolation on

racers. From a systematic review and meta-analysis standpoint, [Mehra, Boyd, and Ickovics \(2017\)](#) recommended race-stratified analyses for a better understanding of the potential roles of residential segregation in health research.

The research protocol and access to the data for this study was reviewed by the Tennessee Department of Health (TDH) Institutional Review Board (IRB) and was approved on February 1st, 2018 (TDH-IRB 1057486) with continuation approval on August 8th, 2021 (TDH-IRB 2020-0152). The National Institutes of Health – Intramural Research Program IRB – Human Research Protections Program – Office of Human Subjects Research Protections determined that the research protocol for this study did not involve human subjects, and thus was exempt from IRB review (18-NIMHD-00722).

### 2.1. Individual-level variables

Cancer and socio-demographic data at the individual level were obtained from the Tennessee Cancer Registry (TCR), which collects information on all Tennessee residents diagnosed with and/or treated for cancer. The TCR is a gold-certified registry, the highest level of certification by the North American Association of Central Cancer Registries. To avoid missing data in the subsequent multilevel analyses, the inclusion criteria were as follows: patients were diagnosed with a non-invasive or invasive breast cancer (classified by the International Classification of Diseases for Oncology, 3rd Edition [ICD-O-3]) during the period between 2005 and 2014; were female (excluded male); provided a self-reported racial identification (either White [non-Hispanic White] or Black [non-Hispanic Black or African American]); were older than 20 years of age; and had a valid county identifier (county of residence at the time of diagnosis). A total of 54,950 individual-level records (85.5% Whites and 14.5% Blacks) were considered in this study.

Since different types of breast cancer can be classified into two broad categories, non-invasive (Stage 0) and invasive (Stages I through IV), these two categories were used to define the individual-level difference in breast cancer condition (the binary outcome of interest). In addition, age, marital status, and health insurance status were used to quantify the individual-level differences among women diagnosed with breast cancer. A description of the study sample (stratified by race) is shown in [Table 1](#).

### 2.2. County-level variables

Demographic data at the county and census tract levels were obtained from the American Community Survey (ACS) for the periods of 2005–2009 and 2010–2014. These 5-year estimates are based on a larger sample size, and thus more reliable than the 1- and 3-year esti-

**Table 1**  
Descriptions of women diagnosed with breast cancer in Tennessee, 2005–2014.

		White	Black
Number of Women (n)		46,983	7,967
Individual-level Outcome			
Breast Cancer	Non-Invasive	17.87%	17.13%
	Invasive	82.13%	82.87%
Individual-level Characteristics			
Age	Minimum	20	21
	Mean	62.24	58.34
	Median	62	57
	Maximum	105	104
Marital Status	Married/Domestic Partner	54.36%	32.68%
	Single/Never Married	8.79%	26.18%
	Divorced/Separated	9.51%	14.53%
	Widowed/Widower	14.96%	13.83%
	Unknown	12.38%	12.77%
Health Insurance	Private Insurance	46.82%	49.17%
	Public Insurance	47.32%	44.27%
	Not Insured/Self-Pay	0.99%	2.25%
	Unknown	4.88%	4.32%

mates. In adherence with the racial/ethnic categories classified in the ACS, county-level estimates of population counts were used to calculate proportions of Whites and Blacks, and subsequently used as proxy measures of county-level White and Black isolation, respectively. Based on the census-tract-level estimates, formal measures of county-level White and Black isolation were derived from the isolation index of  $P^*$  (Bell, 1954; Lieberman, 1981):

$$P^* = \sum_{i=1}^n \left( \frac{g_i}{G} \times \frac{g_i}{t_i} \right)$$

where  $g_i$  is the population count of group  $G$  in census tract  $i$ ,  $G$  is the population count of group  $G$  in the entire study area, and  $t_i$  is the total population count in census tract  $i$ . After calculating values at the census tract level ( $i = 1, 2, 3, \dots, n$ ), they are summarized at the county level (expressed by the summation symbol,  $\sum$ ). Therefore, for White isolation,  $g_i$  and  $G$  in the equation above were substituted by the population count of non-Hispanic Whites in census tracts and in the entire state of Tennessee, respectively. Similarly, for Black isolation,  $g_i$  and  $G$  in the equation above were substituted by the population count of non-Hispanic Blacks in census tracts and in the entire state of Tennessee, respectively. After simply obtaining proxy measures and deriving formal measures, these four county-level measures were used to quantify the degrees of White and Black isolation across 95 counties in the State of Tennessee.

As a supplemental note, formal measures of county-level racial isolation capture the extent to which a racial group is exposed to (or interacts with) the same racial group within each county. On the other hand, its proxy counterparts describe the amount of a racial group in comparative relation to the whole racial and ethnic groups in each county. Therefore, two types of measures have different meanings.

In addition to the demographic data, income data at the county level were also obtained from the ACS for the periods of 2005–2009 and 2010–2014. Owing to the fact that the ACS estimates are represented by the standard hierarchy of census geographic entities (e.g., defined by census tract, county, and state boundaries), both median household income and median family income are widely used and accepted as important indicators of economic well-being of communities, over and above individuals. Since county-level median household income and county-level median family income had a perfect, positive correlation ( $r = 0.98$ ) in both the 2005–2009 and 2010–2014 ACS estimates, county-level median household income was used to quantify the degree of economic well-being across 95 counties in the State of Tennessee.

To put the following multilevel analyses in perspective, racial and ethnic compositions of Tennessee are shown in Table 2 and descriptive statistics of county-level measures are summarized in Table 3. For a better description of the dissimilarities between proxy and formal measures of county-level racial isolation, Pearson’s correlation matrixes are presented in Table 4.

### 2.3. Multilevel analyses

Individual-level and county-level measures were combined into one

**Table 2**  
Racial and ethnic compositions of Tennessee, 2005–2014.

	2005–2009 <sup>a</sup>	2010–2014 <sup>a</sup>
Total Population	6,158,953	6,451,365
White	76.96%	75.01%
Black	16.41%	16.66%
Hispanic or Latino/a	3.71%	4.80%
Asian	1.33%	1.52%
Others <sup>b</sup>	1.59%	2.02%

<sup>a</sup> Based on the state-level estimates from the American Community Survey.

<sup>b</sup> Includes American Indian and Alaska Native, Native Hawaiian and other Pacific Islander, and some other race.

dataset by the five-digit Federal Information Processing Standards county codes in which the year of diagnosis was matched against the time periods of ACS. Using breast cancer condition (non-invasive versus invasive) as the binary outcome of interest, a series of multilevel logistic regression analyses was conducted to estimate the odds ratios (ORs) and associated 95% confidence intervals (CIs). The results from race-stratified models are shown in Tables 5 and 6.

For easier interpretation of the regression coefficients, age was scaled by dividing by 10 to reflect a 10-year increase in age (instead of the unscaled age that reflects a one-year increase in age). Other individual-level measures were handled as categorical variables. Because residential segregation is a unique and unnatural phenomenon (e.g., Charles, 2003; Massey & Fischer, 2000; Williams, 1999), both proxy and formal measures of county-level White and Black isolation followed highly skewed distributions (Table 3). To ensure a meaningful comparison between different units of measurement, proxy and formal measures were standardized by dividing by its interquartile range (IQR). Note that the IQR is the distance between the first quartile and third quartile. In addition, log transformation ( $\ln$  or  $\log_e$ ) was applied to proxy and formal measures for assessing whether log-transformed measures would improve the association between county-level racial isolation and breast cancer condition. Since a proportion of Whites ( $x$ ) was negatively skewed or left-skewed (interpretable by the negative value of skewness in Table 3), log transformation was applied after subtracting from one (1):  $\log_e(1 - x)$ . Finally, county-level median household income followed a slightly skewed distribution (Table 3). Therefore, to compare counties of low economic well-being (i.e., typical low-income counties) with counties of high economic well-being (i.e., typical high-income counties), county-level median household income was standardized by dividing by its IQR.

In Tables 5 and 6, Models 1, 2, 3, and 4 are shown to compare the associations between county-level racial isolation and breast cancer condition with regard to the IQR increase of proxy measure, the increase of proxy measure on a logarithmic scale, the IQR increase of formal measure, and the increase of formal measure on a logarithmic scale, respectively, holding county-level median household income and individual-level measures constant.

Data management, computation of formal measures of county-level White and Black isolation (derived from the isolation index of  $P^*$ ), scaling of individual-level measure, standardization and log transformation of county-level measures, and multilevel logistic regression analyses were carried out in R version 4.1.3 (R Core Team, 2022). The glmer function in the lme4 package (Bates, Maechler, Bolker, & Walker, 2015) was used for carrying out a series of multilevel logistic regression analyses.

### 3. Results

Among patients diagnosed with breast cancer in the State of Tennessee between 2005 and 2014, the prevalence of non-invasive and invasive breast cancer (approximately 17.5% and 82.5%, respectively) were very similar between White and Black women (Table 1). While the percentage of married/domestic partner was higher and single/never married was lower among White women than Black women, other individual-level characteristics were also very similar to one another. However, in relation to the racial and ethnic makeup of Tennessee, where Whites and Blacks accounted for approximately 76.0% and 16.5%, respectively, of Tennessee’s population (Table 2), the residential distributions of two racial groups were quite different from one another (partly interpretable by Table 3). Noticeably, few counties had no Black residents and most counties had only a very small number of Black residents. Also, formal measures, particularly Black isolation, were more skewed and heavy-tailed than the proxy measures (interpretable by the larger values of skewness and kurtosis, respectively, in Table 3).

In reflection to the dissimilar shapes of distributions (Table 3), proxy and formal measures of county-level racial isolation had weak or

**Table 3**  
Descriptive statistics of county-level measures in Tennessee, 2005–2014.

2005–2009	Minimum	First Quartile	Median	Third Quartile	Maximum	Skewness	Kurtosis
Proportion of Whites	0.413	0.861	0.916	0.950	0.991	−2.086	4.451
White Isolation	0.001	0.003	0.005	0.009	0.067	2.998	9.713
Proportion of Blacks	0.000	0.016	0.027	0.077	0.505	2.426	5.758
Black Isolation	0.000	0.000	0.000	0.000	0.344	8.613	76.488
Median Household Income	23,413	32,389	35,995	41,757	88,358	2.106	8.692

  

2010–2014	Minimum	First Quartile	Median	Third Quartile	Maximum	Skewness	Kurtosis
Proportion of Whites	0.378	0.855	0.909	0.939	0.978	−2.118	4.639
White Isolation	0.001	0.003	0.005	0.009	0.065	2.839	8.815
Proportion of Blacks	0.002	0.014	0.032	0.081	0.523	2.496	6.206
Black Isolation	0.000	0.000	0.000	0.000	0.339	8.639	76.881
Median Household Income	26,528	34,298	38,350	42,126	91,743	2.543	11.836

**Table 4**  
Relationships between proxy and formal measures of county-level racial isolation.

2005–2009	Proportion of Whites	White Isolation	Proportion of Blacks	Black Isolation
Proportion of Whites	1	−0.34	−0.97	−0.52
White Isolation	−0.34	1	0.29	0.54
Proportion of Blacks	−0.97	0.29	1	0.52
Black Isolation	−0.52	0.54	0.52	1

  

2010–2014	Proportion of Whites	White Isolation	Proportion of Blacks	Black Isolation
Proportion of Whites	1	−0.36	−0.96	−0.54
White Isolation	−0.36	1	0.27	0.49
Proportion of Blacks	−0.96	0.27	1	0.53
Black Isolation	−0.54	0.49	0.53	1

moderate correlations (Table 4): proportion of Whites and White isolation was weakly and negatively correlated with one another ( $r = -0.34$  or  $-0.36$ ), and proportion of Blacks and Black isolation was moderately and positively correlated with one another ( $r = 0.52$  or  $0.53$ ). Importantly, Table 4 also show that proxy measures were inverses of each other ( $r = -0.97$  or  $-0.96$ ), whereas formal measures were moderately and positively correlated with each other ( $r = 0.54$  or  $0.49$ ). Given the dissimilarities between proxy and formal measures of county-level racial isolation (Tables 3 and 4), two types of measures were incorporated separately into a series of multilevel logistic regression analyses to compare its associations with breast cancer condition in race-stratified models (Tables 5 and 6).

As shown in Table 5, neither standardized or log-transformed proxy measure of county-level White isolation (i.e., a proportion of Whites) nor standardized county-level median household income was associated with breast cancer condition among White women (Models 1 and 2). While standardized county-level median household income remained unassociated, both standardized and log-transformed formal measure of county-level White isolation (derived from the isolation index of  $P^*$ ) were associated with lower odds of having an invasive breast cancer among White women (Models 3 and 4). Put differently, the odds of having an invasive breast cancer among White women was 0.04% lower for an IQR increase in county-level White isolation (OR: 0.96, 95% CI: 0.94–0.98; Model 3) and was 0.11% lower for an increase in county-level White isolation on a logarithm scale (OR: 0.89, 95% CI: 0.85–0.94; Model 4).

For the individual-level variables, using proxy or formal measure of

county-level White isolation did not affect the point estimates (Table 5). Among White women, a 10-year increase in age was associated with higher odds of having an invasive breast cancer (OR: 1.03, 95% CI: 1.01–1.05). In addition, compared with White women who were married or living with a domestic partner, those who were single or never married (OR: 1.10, 95% CI: 1.01–1.20), were divorced or separated (OR: 1.26, 95% CI: 1.15–1.37), or were widowed (OR: 1.29, 95% CI: 1.19–1.40) had higher odds of having an invasive breast cancer. Moreover, relative to White women with private insurance, those with public insurance (OR: 1.21, 95% CI: 1.14–1.29), without insurance or self-pay (OR: 1.79, 95% CI: 1.35–2.38), or with unknown status (OR: 1.38, 95% CI: 1.21–1.57) had higher odds of having an invasive breast cancer.

Unlike White women, as shown in Table 6, standardized county-level median household income was associated with breast cancer condition among Black women in Models 1 through 4, and neither standardized or log-transformed proxy measure of county-level Black isolation (i.e., a proportion of Blacks) nor standardized or log-transformed formal measure of county-level Black isolation (derived from the isolation index of  $P^*$ ) were associated in these four models. Put differently, the odds of having an invasive breast cancer among Black women was 0.08% lower for an IQR increase in county-level median household income (OR: 0.92, 95% CI: 0.86–0.98; Models 1 through 4).

For the individual-level variables, using proxy or formal measure of county-level Black isolation did not affect the point estimates (Table 6), where negligible differences are primarily due to rounding errors. Compared with Black women who were married or living with a domestic partner, those who were single or never married (OR: 1.31, 95% CI: 1.12–1.53) or were divorced or separated (OR: 1.30, 95% CI: 1.08–1.57) had higher odds of having an invasive breast cancer. Moreover, relative to Black women with private insurance, those with public insurance (OR: 1.33, 95% CI: 1.16–1.53), without insurance or self-pay (OR: 2.16, 95% CI: 1.30–3.59), or with unknown status (OR: 1.80, 95% CI: 1.27–2.57) had higher odds of having an invasive breast cancer. Unlike White women, however, a 10-year increase in age among Black women was not associated with breast cancer condition (and the association was also in the opposite direction).

While both proxy and formal measures of county-level Black isolation showed null associations (Table 6), these could be an important source of information for systematic review in the future. Therefore, the results from race-stratified models are both presented in this study to avoid publication bias.

#### 4. Discussion

The results from a series of multilevel logistic regression analyses (Tables 5 and 6) highlight the essential precautions for the study of county-level racial isolation and breast cancer condition among White and Black women residing in the State of Tennessee, as well as for the study of residential segregation and health in general. By comparing the

**Table 5**  
Associations between proxy and formal measures of county-level white isolation and breast cancer invasiveness among white women.

	Model 1	Model 2	Model 3	Model 4
<b>Fixed Effects</b>	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)
<b>County-level Variables</b>				
Proportion of Whites (IQR)	1.03 (1.00, 1.07)			
Proportion of Whites (log <sup>a</sup> )		0.95 (0.89, 1.02)		
White Isolation (IQR)			0.96 (0.94, 0.98)	
White Isolation (log)				0.89 (0.85, 0.94)
Median Household Income (IQR)	0.97 (0.94, 1.00)	0.98 (0.94, 1.01)	0.99 (0.96, 1.02)	1.00 (0.97, 1.03)
<b>Individual-level Variables</b>				
Age (per 10 years)	1.03 (1.01, 1.05)	1.03 (1.01, 1.05)	1.03 (1.01, 1.06)	1.03 (1.01, 1.06)
Married/Domestic Partner	Ref.	Ref.	Ref.	Ref.
Single/Never Married	1.10 (1.01, 1.20)	1.10 (1.01, 1.20)	1.10 (1.01, 1.20)	1.10 (1.01, 1.20)
Divorced/Separated	1.26 (1.15, 1.37)	1.26 (1.15, 1.37)	1.26 (1.15, 1.37)	1.26 (1.16, 1.37)
Widowed/Widower	1.29 (1.19, 1.40)	1.29 (1.19, 1.40)	1.29 (1.19, 1.40)	1.29 (1.19, 1.40)
Unknown	0.94 (0.86, 1.02)	0.94 (0.86, 1.02)	0.93 (0.85, 1.01)	0.93 (0.85, 1.00)
Private Insurance	Ref.	Ref.	Ref.	Ref.
Public Insurance	1.21 (1.14, 1.29)	1.21 (1.14, 1.29)	1.21 (1.14, 1.28)	1.21 (1.14, 1.28)
Not Insured/Self-Pay	1.79 (1.35, 2.38)	1.79 (1.35, 2.38)	1.79 (1.35, 2.38)	1.79 (1.35, 2.38)
Unknown	1.38 (1.21, 1.57)	1.38 (1.21, 1.57)	1.38 (1.22, 1.58)	1.39 (1.22, 1.58)
<b>Random Effects</b>				
County (Intercept)	Variance (SD) 0.0281 (0.1676)	Variance (SD) 0.0292 (0.1708)	Variance (SD) 0.0255 (0.1595)	Variance (SD) 0.0216 (0.1469)
Year (Intercept)	0.0006 (0.0250)	0.0007 (0.0273)	0.0005 (0.0220)	0.0005 (0.0214)
<b>Goodness-of-Fit Measures</b>				
Akaike's Information Criterion	43776.018	43776.962	43765.562	43761.8
Bayesian Information Criterion	43889.866	43890.81	43879.41	43875.648
Log-Likelihood	-21875.009	-21875.481	-21869.781	-21867.9
Deviance	43750.018	43750.962	43739.562	43735.8
Residual Degrees of Freedom	46970	46970	46970	46970

OR: odds ratio; CI: confidence interval; Ref: reference category; SD: standard deviation; IQR: interquartile range; log: a natural logarithm.

<sup>a</sup> a natural logarithm of (1 – Proportion of Whites).

associations of proxy and formal measures of county-level White isolation with breast cancer condition among White women (Models 1 and 2 versus Models 3 and 4 in Table 5, respectively), the results corroborate Oka and Wong's (2014) argument that using a proxy measure of residential segregation may not yield similar results as the formal measure. Moreover, by comparing the associations between formal measures of county-level racial isolation and breast cancer condition from race-stratified models (Models 3 and 4 in Table 5 versus Models 3 and 4 in Table 6, respectively), the results corroborate Mehra and her colleagues' (2017) argument that a potential role of residential segregation in health may differ by race. Notwithstanding the importance of these analytical implications, there are arguably different ways to interpret

**Table 6**  
Associations between county-level measures of white isolation and breast cancer invasiveness among black women.

	Model 1	Model 2	Model 3	Model 4
<b>Fixed Effects</b>	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)
<b>County-level Variables</b>				
Proportion of Blacks (IQR)	0.99 (0.96, 1.01)			
Proportion of Blacks (log)		0.97 (0.90, 1.05)		
Black Isolation (IQR)			1.00 (1.00, 1.00)	
Black Isolation (log)				0.98 (0.95, 1.00)
Median Household Income (IQR)	0.92 (0.86, 0.98)	0.92 (0.86, 0.98)	0.92 (0.86, 0.98)	0.92 (0.86, 0.98)
<b>Individual-level Variables</b>				
Age (per 10 years)	0.95 (0.90, 1.00)	0.95 (0.90, 1.00)	0.95 (0.90, 1.00)	0.95 (0.90, 1.00)
Married/Domestic Partner	Ref.	Ref.	Ref.	Ref.
Single/Never Married	1.31 (1.12, 1.53)	1.31 (1.12, 1.52)	1.31 (1.12, 1.53)	1.31 (1.12, 1.53)
Divorced/Separated	1.30 (1.08, 1.57)	1.30 (1.08, 1.57)	1.31 (1.08, 1.58)	1.31 (1.09, 1.58)
Widowed/Widower	1.19 (0.97, 1.46)	1.19 (0.97, 1.46)	1.19 (0.97, 1.47)	1.20 (0.97, 1.47)
Unknown	1.07 (0.87, 1.31)	1.07 (0.88, 1.32)	1.06 (0.86, 1.30)	1.05 (0.86, 1.29)
Private Insurance	Ref.	Ref.	Ref.	Ref.
Public Insurance	1.33 (1.15, 1.53)	1.33 (1.16, 1.53)	1.33 (1.15, 1.52)	1.33 (1.16, 1.53)
Not Insured/Self-Pay	2.16 (1.30, 3.59)	2.16 (1.30, 3.60)	2.15 (1.30, 3.58)	2.18 (1.31, 3.62)
Unknown	1.80 (1.27, 2.57)	1.80 (1.26, 2.57)	1.81 (1.27, 2.58)	1.81 (1.27, 2.57)
<b>Random Effects</b>				
County (Intercept)	Variance (SD) 0.0000 (0.0000)	Variance (SD) 0.0000 (0.0000)	Variance (SD) 0.0000 (0.0000)	Variance (SD) 0.0000 (0.0000)
Year (Intercept)	0.0032 (0.0562)	0.0032 (0.0564)	0.0031 (0.0553)	0.0032 (0.0563)
<b>Goodness-of-Fit Measures</b>				
Akaike's Information Criterion	7263.894	7265.008	7261.168	7262.411
Bayesian Information Criterion	7354.674	7355.787	7351.948	7353.191
Log-Likelihood	-3618.947	-3619.504	-3617.584	-3618.206
Deviance	7237.894	7239.008	7235.168	7236.411
Residual Degrees of Freedom	7954	7954	7954	7954

OR: odds ratio; CI: confidence interval; Ref: reference category; SD: standard deviation; IQR: interquartile range; log: a natural logarithm.

the protective association between county-level White isolation and breast cancer condition among White women (Models 3 and 4 in Table 5) and to perceive the null association between county-level Black isolation and breast cancer condition among Black women (Models 3 and 4 in Table 6). Hence, possible viewpoints on these matters are discussed below.

For interpreting the protective association between county-level White isolation and breast cancer condition among White women (Models 3 and 4 in Table 5), one possible factor that may have played a role is the health-promoting effects of social capital on health (Kawachi, Subramanian, & Kim, 2007). Social capital refers to multidimensional features of social organization (i.e., social participation, social network, civic participation, social support, trust, norm of reciprocity, and sense of community) that foster the efficacy of communities, particularly among members of tightly knit communities, by facilitating coordinated

actions to achieve a common purpose or goal. Because higher degrees of White isolation may be viewed as the extent to which Whites are protecting themselves from social interaction with and/or social exposure to Blacks (Williams & Collins, 2001) and different socioeconomic groups (e.g., low income populations), such tight social bonds may have provided material and social resources (Kawachi et al., 2007) accrued to White women. While a systematic review of prospective studies showed no relationship between social capital and cancer (Choi et al., 2014), an earlier review suggested protective effects of social capital on adverse health outcomes, such as mortality (including suicide), hospitalization, self-rated health, health-related behavior, and depression (Murayama, Fujiwara, & Kawachi, 2012). Therefore, social capital may not directly protect against breast cancer development and progression, but indirectly through tangible and/or intangible resources: (i) dissemination of knowledge about health promotion, (ii) nourishment of healthy behavioral norms or prevention of deviant health-related behaviors through informal social control, (iii) improvement of access to local services and amenities, and (iv) enrichment of psychosocial processes that provide effective support, build self-esteem, and foster mutual respect (Kawachi & Berkman, 2000, 2014). However, further examinations are needed to understand how a specific dimension (or a combination of dimensions) of social capital operate(s) to nurture what types of health-promoting resources, and to elucidate plausible pathways through which those contextual factors shape the development and progression of invasive and non-invasive breast cancer among White women. To this end, both quantitative and qualitative research, as well as mixed-methods research, are needed in future studies.

With respect to the null association between county-level Black isolation and breast cancer condition among Black women (Models 3 and 4 in Table 6), the main reason behind this may be attributed to the racial/ethnic makeup of urban, suburban, and rural areas (Cromartie, 2018; Johnson, 2012). While the rural population change between 1930 and 2010 in the US showed a complex interplay between natural increase (i.e., births minus deaths) and net migration (i.e., in-migrants minus out-migrants), rural America has been significantly less racially and ethnically diverse than urban areas (Johnson, 2012). Based on the demographic analysis of the past decade, for example, nearly 80% of the rural population was Whites, and Blacks accounted only 8% of the rural population (Cromartie, 2018). In terms of regional variations, the 2010 US Census showed that most counties were predominantly Whites, and counties with relatively large concentrations of Blacks were observed only in rural areas of the Southeast region (Johnson, 2012). Although the State of Tennessee lies in the northern part of the Southeast region, the vast majority (a little over 90%) of land area is rural (Tennessee Department of Health, 2018) and is mostly occupied by Whites. In other words, a dozen of cities with a population size of more than 50,000 (Tennessee Department of Health, 2019) are home to a disproportionately large number of Blacks. Because most Blacks are residents of major cities and their peripheries in Tennessee, only a handful of counties contained so-called "African-American neighborhoods" and the rest of the counties contained none or close to none of those neighborhoods. These, in turn, led to the extremely positive skewness of Black isolation (Table 3). Grounded in the fact that conceptual and theoretical studies on residential segregation and race (e.g., Charles, 2003; Massey & Fischer, 2000) and its health implications among Blacks (e.g., Williams, 1999; Williams & Collins, 2001) have been postulated and/or formulated from urban studies (including the fields of sociology, geography, economics, and political science), an effective and meaningful analysis may be to examine an association between Black isolation and breast cancer condition among Black women at the neighborhood (i.e., census tract) level, and perhaps within a city or a metropolitan area. In doing so, a formal measure of Black isolation needs to be derived from the isolation index of  $SI_i$  (Oka & Wong, 2019), which is a modified and refined version of the isolation index of  $P^*$  (Bell, 1954; Lieberman, 1981), designed for capturing variations at the neighborhood (i.e., census tract) level.

Before exploring aforementioned future studies and disseminating the results of this study, however, there are three major limitations that must be recognized. First, individual-level data on socio-demographic characteristics in the TCR did not have information on socioeconomic status (SES), which applies to most (if not all) of the cancer registries in the US. While SES is a multidimensional construct encompassing economic resources, power, and prestige (which is closely associated with wealth that reflect income level, accumulated economic assets, occupational status, and educational attainment, among others), income level and educational attainment are two common and uninterchangeable measures of SES used as individual-level variables in US studies (Braveman et al., 2001, 2005). By analyzing data from a statewide postpartum survey in California, Braveman et al. (2001) emphasized and warned that excluding income level and/or educational attainment in regression analyses could influence the magnitude of association between an explanatory variable of interest and an outcome of interest. Therefore, the results from a series of multilevel logistic regression analyses (Tables 5 and 6) are subject to overestimation and/or underestimation bias due to the lack of individual-level variables pertaining to income and education, as well as other dimensions of SES and unobserved factors.

Second, formal measures of county-level White and Black isolation derived from the isolation index of  $P^*$  (Bell, 1954; Lieberman, 1981) are inherently influenced by the quality of ACS data. While the ACS is an on-going, nationwide survey conducted every year by the US Census Bureau since 2005, it no longer collects detailed information on demographic, economic, housing and social characteristics through the "long form," which had been used in previous decennial censuses. Because of the changes made in the ACS in which the "long form" was discontinued and replaced by the "short form," estimates for census tracts (but, more so for block groups) are not as reliable (with relatively large margin of errors) as the past decennial censuses (Herman, 2008). This is primarily due to the smaller sample sizes (Herman, 2008) and is particularly evident in sparsely populated areas (Bazuin & Fraser, 2013; Wong & Sun, 2013). However, how the quality of ACS estimates may affect the measurement of residential segregation in general has not been investigated. Therefore, a curtain of uncertainty will be casted over the formal measures of county-level White and Black isolation because the isolation index of  $P^*$  (Bell, 1954; Lieberman, 1981) depends on the quality of census tract estimates.

Third, statistical analyses of place effects on health are bound to be influenced by the presence of spatial autocorrelation. Spatial autocorrelation refers to the dependencies that exist among observations attributed to a clustering of similar or dissimilar values across geographic space (Griffith, 1992). The presence of spatial autocorrelation in a regression analysis causes residuals to vary systematically over space, and thus violates one of the key assumptions that residuals are independent and identically distributed; the violation of independence has been known to inflate R-square values, to deflate standard errors, and to overestimate  $t$ -tests (LeSage, 1997; Martin, 1974). While commonly used multilevel regression models (Gelman & Hill, 2007; Hox, 2010; Raudenbush & Bryk, 2002; Snijders & Bosker, 2012) are capable of accounting for within-group (e.g., within-county) dependencies, these models are incapable of accounting for between-group (e.g., between-county) dependencies. In order to account for spatial autocorrelation in the study of residential segregation and health, Oka and Wong (2014) suggested using (Bayesian) generalized additive mixed models (GAMMs) (Wood, 2006), but Oka and Wong (2016) also suggested using (Bayesian) generalized geospatial mixed models (GGAMMs) (Fahrmeir, Kneib, & Lang, 2004). Given the intricacy of GAMMs and GGAMMs, however, Oka and Wong (2014, 2016) recommended using such sophisticated models only for researchers who have an extensive knowledge of modeling complex spatially multileveled data structures. Therefore, a use of GAMMs and/or GGAMMs needs to be explored in future studies with much caution.

## 5. Conclusion

This study contributes toward a better understanding of county-level racial isolation that may or may not, depending on women's race, explain a part of the difference between patients with invasive and non-invasive breast cancer in the State of Tennessee. The results from a series of multilevel logistic regression analyses (Tables 5 and 6) suggest that a higher degree of county-level White isolation may have a protective effect on White women (Models 3 and 4 in Table 5), but a degree of county-level Black isolation may not matter for Black women (Models 3 and 4 in Table 6). More importantly, the results corroborate Oka and Wong's (2014) argument that using a proxy and formal measure of racial isolation may yield different results (Models 1 and 2 versus Models 3 and 4 in Table 5, respectively), and Mehra and her colleagues' (2017) argument that race-stratified analyses would be helpful for uncovering a differential effect of White and Black isolation in cancer research (Models 3 and 4 in Tables 5 and 6). Despite the usefulness of these findings, however, future studies are needed to elucidate how a higher degree of White isolation shapes the difference between White women with invasive and non-invasive breast cancer, and whether a neighborhood-level (i.e., census-tract-level) Black isolation plays a role in shaping the difference between Black women with invasive and non-invasive breast cancer. Such studies are needed in the State of Tennessee and in other states.

In order to gain a better grasp on the roles of racial isolation in cancer research, however, an ideal research study design is not only to examine a specific outcome of interest (e.g., a type of cancer or a cancer-related behavior), but also to separately examine two or more outcomes of interest (e.g., multiple types of cancer and/or multiple cancer-related behaviors). As partly tackled in this study (Tables 3–6), a key component in such research endeavors is to differentiate and/or distinguish the effects of formal measures from its proxy counterparts. Needless to say, this line of inquiry calls for a thorough understanding of the theoretical and methodological foundations of residential segregation established by demographers, geographers, and sociologists; see Reardon and O'Sullivan (2004) and Johnston et al. (2007) for informative review, and Oka and Wong (2014, 2019) for useful critiques on a use of proxy measures. By embracing a transdisciplinary approach (Rosenfield, 1992) in cancer research, collective insights into the adverse, null, or protective effects of racial isolation in a certain geographic location as well as across different geographic locations are likely to set forth a more rounded view of cancer risk, development, and detection.

## Ethical statement

We have reviewed the Ethics in Publishing and Ethical Guidelines for Journal Publication documents and confirm that we have abided by all ethical guidelines in the production of this manuscript. We have no competing interests or financial interests to disclose.

## Author statement

Masayoshi Oka: Conceptualization, Data Curation, Methodology, Formal Analysis, Writing – Original Draft, Writing – Review & Editing, Visualization. Faustine Williams: Data Curation, Writing – Review & Editing, Project Administration. Martine Whiteside: Data Curation, Writing – Review & Editing, Supervision.

## Declaration of competing interest

None.

## Acknowledgments

Faustine Williams acknowledges support from the Division of Intramural Research, National Institute on Minority Health and Health

Disparities, National Institutes of Health (ZIA MD000015). Martine Whiteside acknowledges support from the Centers for Disease Control & Prevention (Award # 6 NU58DP006307) for the implementation of a population-based cancer registry program in the Office of Cancer Surveillance, Tennessee Department of Health.

## References

- Bates, D., Maechler, M., Bolker, B. M., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Bazuin, J. T., & Fraser, J. C. (2013). How the ACS gets it wrong: The story of the American community survey and a small, inner city neighborhood. *Applied Geography*, 45, 292–302.
- Bell, W. (1954). A probability model for the measurement of ecological segregation. *Social Forces*, 32(4), 357–364.
- Braveman, P. A., Cubbin, C., Egerter, S., Chideya, S., Marchi, K. S., Metzler, M., et al. (2005). Socioeconomic status in health research: One size does not fit all. *Journal of the American Medical Association*, 294(22), 2879–2888.
- Braveman, P. A., Cubbin, C., Marchi, K. S., Egerter, S., & Chavez, G. (2001). Measuring socioeconomic status/position in studies of racial/ethnic disparities: Maternal and infant health. *Public Health Reports*, 116(5), 449–463.
- Charles, C. Z. (2003). The dynamics of racial residential segregation. *Annual Review of Sociology*, 29, 167–207.
- Choi, M., Mesa-Frias, M., Nüesch, E., Hargreaves, J., Prieto-Merino, D., Bowling, A., et al. (2014). Social capital, mortality, cardiovascular events and cancer: A systematic review of prospective studies. *International Journal of Epidemiology*, 43(6), 1895–1920.
- Core Team, R. (2022). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Cromartie, J. (2018). *Rural America at a glance* (2018 edition). U.S. Department of Agriculture, Economic Research Service.
- Fahrmeir, L., Kneib, T., & Lang, S. (2004). Penalized structured additive regression for space-time data: A Bayesian perspective. *Statistica Sinica*, 14, 731–761.
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York, NY: Cambridge University Press.
- Griffith, D. A. (1992). What is spatial autocorrelation? Reflections on the past 25 years of spatial statistics. *L'Espece Géographique*, 21(3), 265–280.
- Herman, E. (2008). The American community survey: An introduction to the basics. *Government Information Quarterly*, 25(3), 504–519.
- Hox, J. (2010). *Multilevel analysis: Techniques and applications*. New York, NY: Routledge.
- Johnson, K. M. (2012). *Rural demographic change in the new century: Slower growth, increased diversity*. Durham, NH: Carsey Institute, University of New Hampshire.
- Johnston, R., Poulsen, M., & Forrest, J. (2007). Ethnic and racial segregation in U.S. Metropolitan areas, 1980–2000: The dimensions of segregation revisited. *Urban Affairs Review*, 42(4), 479–504.
- Kawachi, I., & Berkman, L. F. (2000). Social cohesion, social capital, and health. In L. F. Berkman, I. Kawachi (Eds.), *Social epidemiology* (pp. 174–190). New York: Oxford University Press.
- Kawachi, I., & Berkman, L. F. (2014). Social capital, social cohesion, and health. In L. F. Berkman, I. Kawachi, & M. M. Glymour (Eds.), *Social epidemiology* (pp. 290–319). New York: Oxford University Press.
- Kawachi, I., Subramanian, S. V., & Kim, D. (2007). *Social capital and health*. New York, NY: Springer.
- Kramer, M. R., & Hogue, C. R. (2009). Is segregation bad for your health? *Epidemiologic Reviews*, 31(1), 178–194.
- Landrine, H., & Corral, I. (2009). Separate and unequal: Residential segregation and black health disparities. *Ethnicity & Disease*, 19(2), 179–184.
- Landrine, H., Corral, I., Lee, J. G. L., Efrid, J. T., Hall, M. B., & Bess, J. J. (2017). Residential segregation and racial cancer disparities: A systematic review. *Journal of Racial and Ethnic Health Disparities*, 4(6), 1195–1205.
- LeSage, J. P. (1997). Regression analysis of spatial data. *Journal of Regional Analysis and Policy*, 27(2), 83–94.
- Lieberman, S. (1981). An asymmetrical approach to segregation. In C. Peach, V. Robinson, & S. Smith (Eds.), *Ethnic segregation in cities* (pp. 61–82). London: Croom-Helm.
- Martin, R. L. (1974). On spatial dependence, bias and the use of first spatial differences in regression analysis. *Area*, 6(3), 185–194.
- Massey, D. S., & Denton, N. A. (1988). The dimensions of residential segregation. *Social Forces*, 67(2), 281–315.
- Massey, D. S., & Fischer, M. J. (2000). How segregation concentrates poverty. *Ethnic and Racial Studies*, 23(4), 670–691.
- Massey, D. S., White, M. J., & Phua, V.-C. (1996). The dimensions of segregation revisited. *Sociological Methods & Research*, 25(2), 172–206.
- Mehra, R., Boyd, L. M., & Ickovics, J. R. (2017). Racial residential segregation and adverse birth outcomes: A systematic review and meta-analysis. *Social Science & Medicine*, 191, 237–250.
- Morrill, R. L. (1991). On the measure of geographic segregation. *Geography Research Forum*, 11, 25–36.
- Murayama, H., Fujiwara, Y., & Kawachi, I. (2012). Social capital and health: A review of prospective multilevel studies. *Journal of Epidemiology*, 22(3), 179–187.
- Oka, M., & Wong, D. W. S. (2014). Capturing the two dimensions of residential segregation at the neighborhood level for health research. *Frontiers in Public Health*, 2, 118.

- Oka, M., & Wong, D. W. S. (2016). Spatializing area-based measures of neighborhood characteristics for multilevel regression analyses: An areal median filtering approach. *Journal of Urban Health*, 93(3), 551–571.
- Oka, M., & Wong, D. W. S. (2019). Segregation: A multi-contextual and multi-faceted phenomenon in stratified societies. In T. Schwanen, & R. van Kempen (Eds.), *Handbook of urban geography* (pp. 255–280). Cheltenham, UK: Edward Elgar Publishing.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models in social and behavioral research: Applications and data analysis methods*. Newbury Park, CA: Sage Publications.
- Reardon, S. F., & Firebaugh, G. (2002). Measures of multigroup segregation. *Sociological Methodology*, 32(1), 33–67.
- Reardon, S. F., & O'Sullivan, D. (2004). Measures of spatial segregation. *Sociological Methodology*, 34(1), 121–162.
- Rosenfield, P. L. (1992). The potential of transdisciplinary research for sustaining and extending linkages between the health and social sciences. *Social Science & Medicine*, 35(11), 1343–1357.
- Shannon, C. E. (1948a). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379–423.
- Shannon, C. E. (1948b). A mathematical theory of communication. *The Bell System Technical Journal*, 27(4), 623–656.
- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: SAGE Publications.
- Tennessee Department of Health. (2018). Rural areas. <https://www.tn.gov/health/cedep/environmental/healthy-places/healthy-places/land-use/lu/rural-areas.html>.
- Tennessee Department of Health. (2019). Urban areas. <https://www.tn.gov/health/cedep/environmental/healthy-places/healthy-places/land-use/lu/urban-areas.html>.
- Theil, H. (1972). *Statistical decomposition analysis: With applications in the social and administrative sciences*. New York, NY: American Elsevier Publishing Company.
- White, M. J. (1986). Segregation and diversity measures in population distribution. *Population Index*, 52(2), 198–221.
- White, K., & Borrell, L. N. (2011). Racial/ethnic residential segregation: Framing the context of health risk and health disparities. *Health & Place*, 17(2), 438–448.
- Williams, D. R. (1999). Race, socioeconomic status, and health: The added effects of racism and discrimination. *Annals of the New York Academy of Sciences*, 896, 173–188.
- Williams, D. R., & Collins, C. (2001). Racial residential segregation: A fundamental cause of racial disparities in health. *Public Health Reports*, 116(5), 404–416.
- Wong, D. W. S., & Sun, M. (2013). Handling data quality information of survey data in GIS: A case of using the American community survey data. *Spatial Demography*, 1(1), 3–16.
- Wood, S. N. (2006). *Generalized additive models: An introduction with R*. Boca Raton, FL: Chapman and Hall/CRC Press.