8-7-2018

Evaluating Racial and Geospatial Disparities and Contextual Factors in Triple-Negative Breast Cancer among Women with Breast Cancer

Lia Scott

Follow this and additional works at: https://scholarworks.gsu.edu/sph_diss

Recommended Citation
doi: https://doi.org/10.57709/12489985

This Dissertation is brought to you for free and open access by the School of Public Health at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Public Health Dissertations by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.
EVALUATING RACIAL AND GEOSPATIAL DISPARITIES AND CONTEXTUAL FACTORS IN TRIPLE-NEGATIVE BREAST CANCER AMONG WOMEN WITH BREAST CANCER

by

Lia Cenni Barnar Scott
B.S. Elizabeth City State University
M.P.H. Georgia State University

A Dissertation Submitted to the Graduate Faculty of Georgia State University in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY IN PUBLIC HEALTH (EPIDEMIOLOGY)

ATLANTA, GEORGIA
30303
RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER

APPROVAL PAGE

Evaluating Racial and Geospatial Disparities and Contextual Factors in Triple-Negative Breast Cancer among Women with Breast Cancer

by

Lia Cenni Barnar Scott

Approved:

Lee Rivers Mobley, PhD
Committee Chair

Dora Il’yasova, PhD
Committee Member

Tzy-Mey Kuo, PhD
Committee Member

June 25, 2018
Date
In presenting this dissertation as a partial fulfillment of the requirements for an advanced degree from Georgia State University, I agree that the Library of the University shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to quote from, to copy from, or to publish this dissertation may be granted by the author or, in his/her absence, by the professor under whose direction it was written, or in his/her absence, by the Associate Dean, School of Public Health. Such quoting, copying, or publishing must be solely for scholarly purposes and will not involve potential financial gain. It is understood that any copying from or publication of this dissertation which involves potential financial gain will not be allowed without written permission of the author.

Lia Scott

Signature of Author
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title Page</td>
<td>I</td>
</tr>
<tr>
<td>Approval Page</td>
<td>II</td>
</tr>
<tr>
<td>Author’s Statement</td>
<td>III</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>IV</td>
</tr>
<tr>
<td>List of Tables</td>
<td>V</td>
</tr>
<tr>
<td>List of Figures</td>
<td>VI</td>
</tr>
<tr>
<td>Abstract</td>
<td>1</td>
</tr>
<tr>
<td>Chapter 1. Literature Review and Statement of Purpose</td>
<td>3</td>
</tr>
<tr>
<td>Chapter 2. Descriptive Analysis of Black/White Disparities in Triple Negative Breast Cancer for the US – A Population Based Study from the USCS database</td>
<td>29</td>
</tr>
<tr>
<td>Chapter 3. Examination of Geospatial Patterns in Triple-Negative Breast Cancer and Factors across the United States</td>
<td>47</td>
</tr>
<tr>
<td>Chapter 4. Multi-level Analysis of Person-, County- and State-Level Contributors to Triple-Negative Breast Cancer Diagnosis among Women in the United States</td>
<td>76</td>
</tr>
<tr>
<td>Chapter 5. Dissertation Summary and Future Directions in Research</td>
<td>103</td>
</tr>
<tr>
<td>References</td>
<td>111</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1.1 Descriptive Statistics of Study Population, n=1151724

Table 1.2 Differences in Frequencies of Age, Race, and Stage for Other Breast Cancers and Triple Negative Breast Cancer

Table 1.3 Association between TNBC diagnosis and Race/Ethnicity, Age and Stage at Diagnosis

Table 1.4, Unadjusted and Adjusted Logistic Regression Results for Black/White Subset Analysis, n=973293

Table 2.1, Descriptive Statistics for County Level Variables

Table 2.2 Results of Fitting Multiple Regression Models for Predictors of Breast Cancer Rates (n=2430)

Table 2.3 Results of Fitting Multiple Regression Models for Predictors of Triple Negative Breast Cancer Rates (n=2430)

Table 3.1 Descriptive Statistics for County and State Level Variables

Table 3.2, Mixed Modeling Results for Three-Level Binary Outcome Models to Predict Triple-Negative Diagnosis among Females with Breast Cancer
LIST OF FIGURES

Figure 1 Socioecological model on the Cancer Continuum (Gomez et al., 2015) 19
Figure 1.1 Interaction Effects of Stage at Diagnosis on Age at Diagnosis by Race 43
Figure 2.1 Descriptive Map of 2010-2014 overall breast cancer rates defined by quartile breaks 64
Figure 2.2 Descriptive Map of 2010-2014 triple-negative breast cancer rates defined by quartile breaks 64
Figure 2.3 LISA Cluster Map of 2010-2014 overall breast cancer rates 65
Figure 2.4 LISA Cluster Map of 2010-2014 triple-negative breast cancer rates 65
Figure 2.5 Coincident LISA Clusters Centers 2010-2014 for BC and TNBC rates 66
Figure 2.6 Coincident LISA Clusters 2010-2014 for BC and TNBC rates 66
Figure 3.1 Socioecological model on the Cancer Continuum (Gomez et al., 2015) 90
RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER

ABSTRACT OF THE DISSERTATION
Evaluating Racial and Geospatial Disparities and Contextual Factors in Triple-Negative Breast Cancer among Women with Breast Cancer

by

Lia Cenni Barnar Scott

The objective of this study is to examine racial and geospatial disparities in triple-negative breast cancer diagnosis. Breast cancer, in general, carries an enormous public health burden. Triple-negative breast cancer has greater morbidity and mortality, presenting approximately 9% of all breast cancer diagnoses, in this study. This type of breast cancer has been significantly associated with younger age, African American race, later stage diagnosis, lower socioeconomic status and worse survival. The proposed study will be the first of its kind to use data from the United States Cancer Statistics database which includes combined cancer incidence data from the Centers for Disease Control and Prevention’s (CDC) National Program of Cancer Registries (NPCR) and the Surveillance, Epidemiology, and End Results (SEER) Program covering 99% of the population in comparison to 28% with SEER data alone. This study evaluates individual, social and physical environmental factors associated with disparate rates of diagnosis. Predictors of interest include person level-predictors (race, age, and stage of diagnosis), county-level predictors (residential segregation, social capital and socioeconomic climate), and state-level predictors (breast cancer screening mandates, state-level underinsured rates and state-specific restrictions on nurse practitioner or physician assistant scope of practice). Descriptive epidemiologic analysis allowed us to compare incidence of triple negative breast cancer across race and age groups at the individual level. This study confirmed that Non-Hispanic black women consistently had approximately twice the odds of diagnosis of triple negative breast cancer given breast cancer diagnosis, when compared to Non-Hispanic white women. Younger age and late stage diagnosis
RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER

also conferred higher odds. Exploratory spatial data analysis was used to create descriptive maps
and evaluate patterns of geospatial clustering and underlying community characteristics. This
study found distinct patterns of breast cancer and triple negative breast cancer rates at the county
level. It found 159 counties where breast cancer rates and triple negative breast cancer rates were
concurrently high and 97 counties where breast cancer rates were low, but triple negative breast
cancer rates were high. Spatial regression techniques demonstrated that residential segregation
was consistently associated with both breast and triple negative breast cancer rates. Isolation was
found to be detrimental while diversity was advantageous. Multilevel modeling allowed the
exploration of predictors of triple-negative breast cancer diagnosis at the individual level.
Consistently, race, age and late stage diagnosis conferred higher odds of diagnosis with triple
negative breast cancer, given breast cancer diagnosis. Residential segregation measures were
consistently associated, with isolation conferring higher odds and diversity conferring lower odds
of diagnosis. The results of these studies potentially inform policy at actionable geographic
levels and add valuable information to cancer health disparities research. Additionally, they
provide insight that there is still a need to explore what factors may be driving racial and
geospatial disparities in triple negative breast cancer in the United States.
Chapter 1

Literature Review and Statement of Purpose
Epidemiology of Triple-Negative Breast Cancer

Carcinoma of the breast, or breast cancer, is the most common type of noncutaneous cancer among women. It is expected that there will be 63,960 in situ cases, 266,120 invasive cases, and 40,920 deaths in women in 2018. The probability of developing invasive breast cancer increases with age and is 12.4% or 1 in 8 for a lifetime. (Siegel, Miller, & Jemal, 2018). When we examine the five leading causes of cancer death in 2015, among females, breast cancer is the second only behind lung and bronchus. However, for age 20 to 59, it is the first leading cause of cancer death (Siegel et al., 2018). Approximately 80% of breast cancers are invasive with up to 21 distinct histological subtypes and at least 4 different molecular subtypes (American Cancer Society, 2017).

Numerous biologic subtypes of breast cancer demonstrate that it is a heterogeneous disease (Carey et al., 2006). These subtypes are based on gene expression patterns that include 496 genes (Perou et al., 2000). There are five intrinsic subtypes recognized. Two estrogen receptor (ER)-positive types include luminal A (ER+/progesterone receptor (PR)-positive and human epidermal growth factor receptor 2(HER2)-negative) and luminal B (ER+/PR+/HER2+). Three ER-negative types include ER-/PR-/HER2+, basal-like (ER-/PR-/HER2-/cytokeratin 5/6 (CK5/6) +), and unclassified ‘normal-like’ (negative for all markers) (Yang, Pfeiffer, et al., 2007; Yang, Sherman, et al., 2007). These subtypes differ markedly in prognosis and therapeutic targets (Sorlie et al., 2001). Luminal A tumors have the most favorable clinical features, followed by luminal B (Yang et al., 2007). ER positive tumors respond to endocrine therapy such as antiestrogen administration or ovarian suppression. HER2 positivity provides bases for targeted therapy with monoclonal antibody against HER2 (Bauer et al., 2007) Basal-like tumors have poor clinical features and survival, due to the negative ER and HER2 status. ER and PR
negative tumors account for approximately 20% of breast cancers with known receptor status and include the most clinically aggressive tumors (Yang et al., 2007). Triple negative breast cancer (TNBC) accounts for approximately 12-15% of all breast cancer cases and approximately 75-90% are deemed basal-like (Dolle et al., 2009; American Cancer Society, 2017). TNBC is associated with aggressive histology, poorer prognosis, shorter survival, and unresponsiveness to usual hormone and HER2 immunotherapy (Bauer, Brown, Cress, Parise, & Caggiano, 2007). This is important to note as basal-like tumors have a characteristic morphology that includes high proliferative rate, central necrosis, and a pushing border (when the edge of the tumor appears to be pushing into normal tissue) and have been associated with aggressive histology, unresponsive to hormone therapy, poor prognosis, and BRCA-1 gene mutation (Dolle et al., 2009; Kreike et al., 2007; Livasy et al., 2006; Foulkes et al., 2003; Perou et al., 2000; Sorlie et al., 2001; Sorlie et al., 2003; Foulkes et al., 2010).

Triple negative breast cancers are more likely than any other breast cancer type to metastasize to viscera, specifically the lungs and the brain (Foulkes et al., 2010). There is approximately a 40% chance of first distant recurrence in the lungs for metastatic triple negative breast cancer, while there is only a 20% chance for non-triple negative breast cancer (Foulkes et al., 2010). A study conducted using the California cancer registry found that the triple negative phenotype was statistically significantly associated with younger age, Non-Hispanic black race/ethnicity, later stage diagnosis, lower SES and shortened survival (Bauer et al., 2007). Non-Hispanic black women also have statistically significant earlier age at diagnosis, higher proportion of high grade tumors, and a higher proportion of triple-negative breast cancers suggesting that breast cancer in Non-Hispanic black women is biologically different (Chu, Lamar, & Freeman, 2003; Newman, 2005). These disparate outcomes and lack of knowledge of
RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER

etiology of disease provide reason as to why research must shift focus from treatment to identifying risk factors, albeit environmental or genetic, that exacerbate disparities in breast cancer diagnosis to develop and implement more efficacious population-based prevention strategies.

Individual Risk Factors

Race - Disparities between Non-Hispanic black and Non-Hispanic white women

Racial disparities exist within breast cancer diagnoses, particularly between non-Hispanic white and non-Hispanic black populations. Incidence trends from 2005 to 2014 demonstrate that invasive breast cancer rates were stable for non-Hispanic white women and increased slightly for non-Hispanic black women. Over the 10-year span, invasive breast cancer rates for Non-Hispanic black women increased approximately 3%. Overall breast cancer mortality rates have declined since 1989. Mortality peaked in 1989 at 33.2 deaths per 100,000 and declined 39% to 20.3 deaths per 100,000 in 2015. This decline has been attributed to early detection and screening. The annual percent decline from 2006 to 2015 was larger for non-Hispanic white women compared to non-Hispanic black women, 1.8% versus 1.5%, respectively (Siegel et al., 2018).

According to SEER data from 18 registries, the lifetime relative risk of all breast cancers from 2013-2015 for Non-Hispanic black women compared to Non-Hispanic white women is 0.90, indicating Non-Hispanic black women are at lower risk for diagnosis, yet the lifetime relative risk for dying of breast cancer is 1.22. When Non-Hispanic black women are diagnosed with breast cancer, there are 22% more likely to die from the disease. Age-adjusted incidence rates from 2011-2015 were 128.6 per 100,000 for non-Hispanic white women and 126.9 per 100,000 for non-Hispanic black women. The age-adjusted mortality rates from breast cancer in
RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER

2011-2015 for Non-Hispanic black women was 28.7 per 100,000 and 20.3 deaths per 100,000 for Non-Hispanic white women (Noone et al., 2018). When examining cases from 2008 to 2014, Non-Hispanic white women had a higher percentage of localized disease as compared to Non-Hispanic black women (63% vs 54%), yet the 5-year relative survival rate for Non-Hispanic black women was lower in terms of localized (99.1% vs 95.4%), regional (86.4% vs 76.6%), and distant stages (28.1% vs.19.7%) (Noone et al., 2018). Survival in Non-Hispanic black women may be worse due to a higher frequency of adverse histologic features (Bauer et al., 2007). Incidence data are from the 18 registries, while mortality data are for the entire United States.

*Lifetime Exposure to Estrogen*

Besides race, research has established multiple individual risk factors associated with breast cancer diagnoses, although the complete etiology of the disease remains unknown. Lifetime exposure to estrogen has been linked to increased risk for breast cancer (Loeffler & Hart, 2014). These factors include early menarche, late menopause, no or fewer children, and receiving exogenous estrogen (Loeffler & Hart, 2014). In addition, women who had children but did not breastfeed are at increased risk compared to those who breastfed (Loeffler & Hart, 2014). High levels of circulating estrogens and androgens have been associated with increased risk of breast cancer in premenopausal women (Key et al., 2013). Parity also plays a role. Having a first child before 35 and a greater number of children is associated with decreased risk for hormone receptor positive breast cancers (Lambertini et al., 2016). On the contrary, there is an increase in hormone receptor negative breast cancer risk that last about 10 years post full-term pregnancy, particularly among women who are older at birth (Albrektsen, Heuch, Hansen, & Kvåle, 2005; Schedin, 2006). Breastfeeding also slightly reduces overall risk of breast cancer if done for at least a year (Faupel-Badger et al., 2012). The risk of breast cancer is reduced by 4%
RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER

for every 12 months of breastfeeding (Collaborative Group on Hormonal Factors in Breast Cancer, 2002). This may be due to the inhibition of menstruation when breastfeeding, thus reducing lifetime number of menstrual cycles, or the structural changes that occur in the breast following lactation and weaning (Britt, Ashworth, & Smalley, 2007; Faupel-Badger et al., 2012). This effect was found to be stronger among triple-negative cases (Faupel-Badger et al., 2012; Islami et al., 2015; Sisti et al., 2016).

Family History

Women with a family history of breast cancer, especially in a first-degree relative have an increased risk of breast cancer diagnosis. When we compare them to women without a family history, their risk of diagnosis is twice as high when there is only one affected first degree female relative, and almost 4 times as high when there are multiple first-degree female relatives with breast cancer (Collaborative Group on Hormonal Factors in Breast Cancer, 2001). Additionally, a family history of ovarian cancer is also associated with higher risk of breast cancer diagnosis (American Cancer Society, 2017). Women who have been diagnosed with breast cancer have a small increased risk of developing a new cancer in the opposite breast (Nichols, De González, Lacey Jr, Rosenberg, & Anderson, 2011). Ductal carcinoma in situ (DCIS) and Lobular Carcinomas in situ (LCIS) are both potential precursors for invasive breast cancer. Women with a history of DCIS are 10 times more likely to be diagnosed with an invasive breast cancer and women with LCIS are 7 to 12 times more likely to develop invasive breast cancer compared to women with either (King et al., 2015; Lopez-Garcia, Geyer, Lacroix-Triki, Marchió, & Reis-Filho, 2010; Morrow, Schnitt, & Norton, 2015).

Genetic Predisposition
The most indicative of these risk factors is the mutation of the \textit{BRCA} genes. \textit{BRCA} genes are tumor suppression genes that repair damaged DNA and prevent cell division in damaged cells. A mutation in either \textit{BRCA-1} or \textit{BRCA-2} increases the risk of breast cancer from 12\% up to 70\% (Kuchenbaecker et al., 2017). These mutations occur at a rate of less than 1\% in the general population but occur slightly more often in certain ethnic or geographically isolated groups (Gabai-Kapara et al., 2014). Another gene that works with \textit{BRCA-2}, \textit{PALB-2}, appears to confer risk that may be as high as \textit{BRCA-2} mutations (Antoniou et al., 2014). Multiple research avenues are being explored to determine what oncogenes or tumor suppressor genes play a role in the various subtypes of breast cancer (Li et al., 2015; McLean et al., 2015; Plíšková, Vondráček, Vojtěšek, Kozubík, & Machala, 2004). There are no known recurring breast cancer gene mutations that have an increased frequency in non-Hispanic black women (Chen et al., 1994; Cross, Harris, & Recht, 2002; Gao, Neuhausen, Cummings, Luce, & Olopade, 1997; Newman et al., 1998; Shen et al., 2000). Additionally, not much is known about the interaction of genes and environmental factors that may be different between non-Hispanic black and non-Hispanic white women (Cross et al., 2002).

\textit{Height}

Height has also been associated with increased breast cancer risk (Green et al., 2011; van den Brandt et al., 2000). A study found that in a European sample, height was an independent risk factor for breast cancer among postmenopausal women, but the relationship was unclear among premenopausal women (van den Brandt et al., 2000). Increased height was associated with increased risk of cancer and cancer death in another European study (Wirén et al., 2014). Height may be reflective of differences in early growth or hormonal factors.

\textit{Obesity}
Obesity plays a different role in breast cancer risk when we compare pre- and post-menopausal women. In a meta-analysis study of premenopausal women, breast cancer risk was 1% lower in overweight women and 26% lower in obese women compared to women of normal weight (Nelson et al., 2012). However, this may be limited to hormone receptor positive tumors. The postmenopausal risk of breast cancer is 1.5 times higher in overweight women and 2 times higher in obese women (La Vecchia, Giordano, Hortobagyi, & Chabner, 2011). Obesity is also a risk factor for type II diabetes which is linked to increases risk of postmenopausal breast cancer (16% increase) (Boyle et al., 2012; Maskarinec et al., 2017; Tsilidis, Kasimis, Lopez, Ntzani, & Ioannidis, 2015). Women who exercise regularly had 10-20% lower risk of breast cancer compared to women who are inactive (Pizot et al., 2016). The reduction is greater with increasing amounts of exercise and vigorous activity, but even less vigorous physical activity, such as walking, is beneficial (Hildebrand, Gapstur, Campbell, Gaudet, & Patel, 2013). This may be because of the effect of physical activity on inflammation, hormones, and energy balance (Neilson, Friedenreich, Brockton, & Millikan, 2009; Pizot et al., 2016).

**Determination of Social Environmental Risk Factors**

The identification of factors that create and exacerbate these disparities would be an ideal outcome of this and future cancer epidemiology studies. Biological risk factors have been identified and are consistently validated, yet disparities still exist. Dating back to 1991, socioeconomic status (SES) has been studied as a factor in cancer incidence. A study examining the association between census tract level income and educational levels as proxies for socioeconomic status and cancer incidence at all sites combined found that the disproportionate distribution of the non-Hispanic black population at lower socioeconomic levels accounts for much of the excess burden among this population (Baquet, Horm, Gibbs, & Greenwald, 1991).
SES variables are not collected with patient characteristics in cancer registries. Thus, it is not surprising that few studies have explored the role of social environmental risk factors in triple negative breast cancer diagnosis, specifically (Chu, Henderson, Ampil, & Li, 2012; Dolle et al., 2009). Collection of additional social environmental risk factors that vary among women can inform both policy and clinical practice. Lacking such person-specific SES data, this study aims to illuminate the role that residential segregation, socioeconomic conditions and other contextual factors in the counties where women live may play in the diagnosis of TNBC. This study looks at the following factors from an ecological standpoint, rather than compositional.

**Residential Segregation**

Residential segregation adversely impacts the health of non-Hispanic black persons (Collins, 1999; Collins & Williams, 1999). Segregation promulgates negative social environments as highly segregated cities often experience higher levels of violent and property crimes (Kramer & Hogue, 2009; Peterson & Krivo, 1993; Velez, Krivo, & Peterson, 2003). Residential segregation impacts access to health-relevant sources. Even after controlling for risk factors, segregation may have a statistically significant effect on health outcomes because of the way it shapes contact patterns and social networks (Acevedo-Garcia, 2000; Acevedo-Garcia, Lochner, Osypuk, & Subramanian, 2003).

There has research conducted using SEER registry data and a sample of 395,671 US-born non-Hispanic black and non-Hispanic white women that found that Jim Crow birthplace was associated with increased odds of estrogen receptor negative breast cancer among non-Hispanic black women with the strongest effect for women born before 1965 (Krieger, Jahn and Waterman, 2017). In terms of breast cancer care, Haas and colleagues (2008) found that non-Hispanic black segregation was a mediator of the Black/White disparity in breast cancer care.
They also found that both non-Hispanic black and non-Hispanic white women who lived in areas of greater non-Hispanic black segregation were less likely to receive adequate breast cancer care.

**Socioeconomic Status**

Socioeconomic position has been linked to breast cancer incidence. Studies show that this construct is positively related to breast cancer incidence overall, but the association varies by race or ethnicity (Bigby & Holmes, 2005; Yost, Perkins, Cohen, Morris, & Wright, 2001). Research demonstrates that the United States is the most unequal in terms of wealth distribution among all developed countries. The top 1 percent of households own 38 to 47 percent of all wealth (Keister & Moller, 2000; Wolff, 1996, 1998). The Gini coefficient is the most commonly used indicator of income inequality, and its use here allows for comparisons with other income inequality studies (Jones-Smith, Gordon-Larsen, Siddiqi, & Popkin, 2011; Nowatzki, 2012). The association between income inequality and mortality has also been established at cross-country and national levels (Chetty et al., 2016; Kaplan, Pamuk, Lynch, Cohen, & Balfour, 1996; McIsaac & Wilkinson, 1997).

Education is often considered a social determinant of health under the umbrella of socioeconomic status. One study found that having low self-reported education was associated with subtypes of estrogen receptor negative and progesterone receptor negative breast cancers (Trivers et al., 2009). Additionally, women in their study with triple negative tumors were more often of lower SES. Low-income women have mammography screening rates that are lower than higher income women (Peek & Han, 2004). Additionally, women with lower education have lower mammography screening rates (Kerner et al., 2001). Women in low-income areas are more likely to present with late stage disease and socioeconomic position has also been association with treatment (Schwartz, Crossley-May, Vigneau, Brown, & Banerjee, 2003).
Low socioeconomic status has been linked to decreased rates of screening, greater probability for late-stage diagnosis, receipt on inadequate and disparate treatment and higher mortality from breast cancer (Bigby & Holmes, 2005; Gerend & Pai, 2008). Regardless of race, poverty is associated with poorer breast cancer outcomes (Gerend & Pai, 2008). Socioeconomic status is often linked to access to care as they often go hand-in-hand. Those with lower socioeconomic status are less likely to have employment stability that provides adequate insurance for care. Additionally, studies have found that the non-Hispanic black population is twice as likely to be uninsured and depend on public insurance compared to the non-Hispanic white population (Thomasson, 2006). Uninsured and underinsured women are less likely to undergo screening, more likely to receive a late stage diagnosis and less likely to survive (Bradley, Given, & Roberts, 2002; Buseman, Byers, Finch, & Jacobellis, 2002; Gerend & Pai, 2008; Gordon, Rundall, & Parker, 1998; Hsia et al., 2000; Roetzheim et al., 1999). Women who reside in disadvantaged communities may be required to travel longer distances with longer wait times to utilize screening and treatment facilities. These factors can cause a major hindrance in regular physician visits (Mandelblatt, Andrews, Kao, Wallace, & Kerner, 2010; Mandelblatt, Andrews, Kerner, Zauber, & Burnett, 1991; Mandelblatt, Yabroff, & Kerner, 1999; Vernon, Vogel, Halabi, & Bondy, 1993).

Higher socioeconomic status confers a higher risk of breast cancer diagnosis, while lower socioeconomic status confers higher risk of triple negative breast cancer diagnosis. In terms of triple negative diagnosis, this type occurs more frequently in younger women and in non-Hispanic black women (Bauer et al., 2007; Dent et al., 2007; Haffty et al., 2006; Harris et al., 2006; Morris et al., 2007; Tischkowitz et al., 2007). Socioeconomic status has been examined as a factor in breast cancer incidence and the National Action Plan on Breast Cancer Workshop
RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER

recommended that it be an integral part of research in breast cancer etiology, since this variable is associated with many other factors in an individual’s life (Dayal, Power, & Chiu, 1982; Gordon, 2003; Krieger, 1990; Moormeier, 1996). Even after these factors have been taken into consideration, racial and geospatial disparities still exist. It is important to explore how these individual and social factors are associated with breast cancer diagnosis, broken down by subtype.

Gaps in the literature and this study’s contributions

Research has primarily been focused on identifying individual risk factors with small sample sizes. There appears to be a consensus that age and race play significant roles in the diagnosis of TNBC, yet with the knowledge we have, disparities still persist. This dissertation aims to validate these small study findings, as it is the first to use a near complete population dataset, the United States Cancer Statistics database (USCS). It takes a step back and evaluates how the social environment may play some role in TNBC diagnosis. Additionally, it allows us to evaluate modifiable risk factors in TNBC, thus is more useful in addressing the disparities between Non-Hispanic black and Non-Hispanic white women. When analyses are done at the local geopolitical level, such as the county, we are able draw conclusions and in turn develop and implement prevention strategies that are more conducive to these well-defined geographic areas.

With the recent, wider availability of the USCS database, more studies are beginning to examine societal and community contributors to breast cancer diagnoses. Some of these have used a spatial analytic perspective (Kuo et al., 2011; Mobley et al., 2017; Mobley & Kuo, 2015a, 2015b; Mobley, Kuo, & Andrews, 2008; Mobley, Kuo, Urato, et al., 2012; Mobley, Kuo, Watson, et al., 2012; Mobley, Scott, Rutherford, & Kuo; Mobley et al., 2017). The few studies that have examined community characteristics have been mainly descriptive (Beyer & Rushton,
RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER

2009; Roche, Skinner, & Weinstein, 2002). Race and residential segregation as representatives for social support has been of particular focus when examining breast cancer disparities (Bradley et al., 2002; Haas et al., 2008a, 2008b; Kuo et al., 2011; Mobley et al., 2017; Warner & Gomez, 2010). The use of the expansive Spatial Impact Factor (SIF) dataset along with other free, public-use population health data allows for more in depth identification and evaluation of additional potential contributors. The SIF database contains a time series of cross sections reflecting multiple community level variables including but not limited to residential segregation, poverty, income inequality, food security and urbanicity. The database is supplemented by additional data from the Area Health Resource File and Pennsylvania State University (Bureau of Health Workforce, 2017; Rupasingha, Goetz, & Freshwater, 2006). These data are drawn primarily from the U.S. Census Bureau data.

Few descriptive studies have explored individual and community (county and state) level contributors to breast cancer outcomes, using multilevel analysis or spatial regression. This study will update previous SEER registry publications by expanding the geographic scope of these studies. The use of the comprehensive USCS database allows for an unprecedented exploration of this health outcome. Additionally, this study will employ statistical methods that allow for inference to accompany the more descriptive studies. Geospatial analysis and statistical and spatial modeling will provide insight to better inform policy and design prevention strategies, particularly due to the inclusion of social environmental risk factors.

Few previous studies have examined social environmental predictors in triple-negative breast cancer. Using the entire USCS registry database would provide enough of these rare cancer cases and a broad spectrum of geospatial heterogeneity, yielding the best possible study design for these disparities. Only recently have USCS database studies been conducted.
RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER

concerning population cancer outcomes associated with state-level health policy factors (Mobley & Kuo, 2015a, 2015b). Previous studies have also found that constructs representative of social support can be influential on both breast cancer screening and stage of diagnosis models (Kuo, Mobley, & Anselin, 2011; Mobley, Kuo, Scott, Rutherford, & Bose, 2017; Mobley, Kuo, Urato, et al., 2012; Mobley, Kuo, Watson, & Gordon Brown, 2012). These studies provide limited evidence that the environment does impact cancer stage at diagnosis. The use of the USCS dataset (U.S. Cancer Statistics Working Group, 2017) allows for an expansion of these studies to fully inform healthcare policy and may potentially reveal contributory factors to typology and staging at diagnosis.

**Methodological Approaches**

This study utilizes a variety of methodological approaches with a comprehensive dataset to address the different aspects of TNBC incidence. The USCS dataset provides the best source of information on population-based cancer incidence for the nation (Henley et al, 2010). Data must meet six USCS publication criteria: 1) case ascertainment is ≥90% complete, 2) ≤5% of cases are ascertained solely on the basis of a death certificate, 3) ≤3% of cases are missing information on sex, 4) ≤3% of cases are missing information on age, 5) ≤5% of cases are missing information on race, and 6) ≥97% of the registry's records passed a set of single-field and inter-field computerized edits that test the validity and logic of data components (Henley et al, 2010). These registries cover approximately 99% of the U.S. population, including 96% of the U.S. non-Hispanic white population, 99% of the U.S. non-Hispanic black population, 89% of the U.S. AI/AN population, 98% of the U.S. API population, and 96% of the U.S. Hispanic population (Henley et al., 2010; U.S. Cancer Statistics Working Group, 2017). This is substantially more comprehensive compared to the SEER Program with covers approximately 28 percent of the
RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER

U.S. population, including 25 percent of non-Hispanic whites, 26 percent of African Americans, 38 percent of Hispanics, 44 percent of American Indians and Alaska Natives, 50 percent of Asians, and 67 percent of Hawaiian/Pacific Islanders (SEER, 2018).

The first study aims to validate the findings of smaller, more localized studies. Traditional epidemiologic analysis is used in this study, such as descriptive statistics, basic inferential statistics and logistic regression. Using only the individual level data, we can determine the odds of diagnosis with TNBC based on race, age and stage of diagnosis for women with breast cancer. This is the first study of its kind to use national data in the exploration of differences in person-level factors. However, the use of these methods can often confound race and place.

The second study aims to supplement the descriptive analysis by evaluating descriptive geospatial patterns of disease at the county-level. Spatial analysis is the primary methodological toolkit used in the proposed study. A variety of geographic information system and spatial analytic approaches have been utilized in previous analyses in the literature. These methods often include simple GIS and mapping, Bayesian image analysis, SaTScan, and generalized linear modeling (Goovaerts, 2010; Gumpertz, Pickle, Miller, & Bell, 2006; MacKinnon et al., 2007; McElroy, Remington, Gangnon, Hariharan, & Andersen, 2006). This study will employ the use exploratory and confirmatory spatial data analysis through descriptive mapping, the Local Moran’s I or LISA statistic and spatial regression. This study focuses on underlying county-level factors and how they vary across extremes in the pattern and is limited to ecological interpretation due to the use of contextual data. The use of geospatial analysis allows us to pinpoint where problems and disparities exist, thus we can better develop prevention programs as well as provide informed, culturally competent treatment. There is great potential to reveal new
RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER

information or confirm existing information on disparities within this outcome due to the various spatial analytic methodologies and datasets proposed for this study.

The third study moves further into explaining how person-, county- and state-level factors can explain TNBC diagnosis at the individual level. To address the fallacies of single-level research, it is important to consider contexts and multi-level phenomena when conducting population health research (Oakes, 2009). Additionally, the multi-level approach allows us to address spatial heterogeneity found in the previous studies from this dissertation, and causality cannot be inferred from this study. Rather, this study aims to inform the literature on potential contextual contributors to disparate rates and explore if disparities persist when controlled for. The socioecological model acknowledges the importance of context and this study aims to elucidate the specific role residential segregation and socioeconomic conditions play in diagnosis (Gomez et al, 2015). Multi-level studies can be used to inform multi-level interventions. We consider the socioecological model of the cancer continuum in this study (Figure 1).
Overall, the proposed study intends to fill several significant gaps in the literature by conducting population-based secondary analyses that will evaluate racial, ethnic and geospatial disparities in triple negative breast cancer. Current evidence provides reasons regarding why research must shift focus from treatment to prevention through the identification of risk factors, albeit environmental or genetic, that exacerbate disparities in breast cancer diagnosis to create and implement more efficacious population-based prevention strategies. The study will use data from United States Cancer Statistics (USCS) database, which includes combined cancer incidence data from the Center for Disease Control and Prevention’s (CDC) National Program of Cancer Registries (NPCR) and the SEER Program. Each data point in the study is considered a case and we expect over 1 million observations in the time period of 2010 to 2014. Case-only studies are useful in understanding the heterogeneity among the cases (Begg & Zhang, 1994;
Trivers et al., 2009). This data covers 99% of the population during 2009-2013 and 92% during 1999-2013 (Richardson, Henley, Miller, Massetti, & Thomas, 2016).

No previous studies have examined disparities and their predictors in triple-negative breast cancer or other aggressive, invasive cancer typologies using the entire USCS database. Both demographic and geographic disparities exist, and these are different constructs that must both be elucidated to inform disparities reduction descriptive analysis is the first national analysis of population-based cancer incidence.
Statement of Purpose

The purpose of this study is to evaluate racial and geographic disparities in triple-negative breast cancer (TNBC) diagnoses by race or ethnicity both within regions and across regions. The primary outcome of interest is TNBC, breast cancer with immuno-histochemical suppression of estrogen-receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2), at first diagnosis. Previous studies have examined disparities but not on the population scale, often focusing on the individual level. Studies also fail to examine multi-level effects, which is what this study will examine. This study will evaluate spatial clustering of TNBC at the county-level and examine environmental predictors of triple-negative breast cancer status using NPCR data available for 45 states, combined with SEER data for the remaining 5 states. This study uses a subset of the data to include states that allow the use of county-level incidence data, have available contextual data, as well as states that code for TNBC. This reduces the state sample size to 39 states. Five states do not provide county-level breast cancer data – Illinois, Kansas, Michigan, Minnesota and Missouri – and four do not code for triple negative data – Connecticut, Iowa, New Mexico, and Utah. Alaska and Hawaii are excluded from analysis due to missing contextual data. This study aims to account for personal, social and environmental factors such as age, race, socioeconomic status, residential segregation and other neighborhood characteristics and their impacts on TNBC outcomes.

The overarching goal of this study is to advance the field of population-based research in breast cancer disparities through innovative statistical techniques. The objective of this study is to address racial and geospatial disparities in triple-negative breast cancer diagnosis and to examine potential predictors of diagnosis. Breast cancer, in general, carries an enormous public health burden and triple-negative breast cancer accounts for 15% of all breast cancer diagnoses.
RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER

Morbidity and mortality burdens are higher with this type of breast cancer and diagnosis has been significantly associated with younger age, African American race, later stage diagnosis, lower socioeconomic status and shortened survival. The proposed study will be the first of its kind to use data from the United States Cancer Statistics database which includes combined cancer incidence data from the Center for Disease Control and Prevention’s (CDC) National Program of Cancer Registries (NPCR) and the Surveillance, Epidemiology, and End Results (SEER) Program covering 99% of the population in comparison to 28% with SEER data alone.

This study will focus on racial disparities between Non-Hispanic black and Non-Hispanic white women and geospatial disparities across the contiguous United States. It will evaluate individual, social and physical environmental factors that contribute to disparate rates of diagnosis and survival. Predictors of interest include, but are not limited to, person level-predictors – race, age, and stage of diagnosis, county-level predictors – residential segregation, social capital and socioeconomic climate, and state-level predictors – breast cancer screening rates, state-specific restrictions on Nurse Practitioner or Physician Assistant scope of practice, and underinsurance or self-insurance policies. Descriptive epidemiologic analysis will allow us to compare incidence of triple negative breast cancer across race and age groups at multiple geographic levels.

Exploratory and confirmatory spatial data analysis will be used to create descriptive maps and evaluate patterns of geospatial clustering and underlying community characteristics. Multilevel modeling with latent variables will allow us to explore predictors of triple-negative breast cancer diagnosis and survival. Results will robustly answer the question of both ‘why?’ and ‘where?’ thus potentially informing policy at actionable geographic levels and adding valuable information to cancer health disparities research as a whole.
RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER

We expect to find evidence of racial and geospatial disparities in TNBC. Additionally, we expect statistical evidence of positive spatial autocorrelation of TNBC rates and that clusters will have distinct community characteristics. Finally, we expect that the use of multi-level and spatial analytic methods will better elucidate the various predictors and pathways of TNBC outcomes and perhaps reveal specific geographic areas in greatest need of prevention or intervention. The usual challenge in tackling this type of analysis is data limitation; however, we largely overcome that by using the USCS database which covers 99% of the cancer population.

Identifying information will be protected due to the secure nature of the Research Data Center and work conducted there, and the de-identification of datasets. Future work will expand upon this spatial analytic research foundation to examine other types of breast or other cancers demonstrate disparate rates, such as colorectal, cervical and prostate cancer.

The use of spatial analysis in a variety of methodologies allows us to robustly answer the questions of ‘Why?’ and ‘Where?’ these cancer diagnoses are occurring. State and county level characteristics will be used to model the spatial heterogeneity of TNBC diagnoses. Mixed modeling allows us to examine multi-level variable predictors and influence on this particular type of cancer diagnoses. There is no literature comparable and the proposed study will help fill several significant gaps in the literature. This data-intensive, information-rich research strategy is the future of applied population health research.

**Study 1. Descriptive Analysis of Black/White Disparities in Triple Negative Breast Cancer for the US – A Population Based Study from the USCS database**

Triple negative breast cancer (TNBC) accounts for approximately 15% of all breast cancer cases, and is associated with aggressive histology, poorer prognosis, shorter survival, and unresponsiveness to usual hormone therapy (Bauer et al., 2007). A study conducted using the
California cancer registry found that the triple negative phenotype was statistically significantly associated with younger age, African American race/ethnicity, later stage diagnosis, lower SES and shortened survival (Field et al., 2012; Field et al., 2005). Non-Hispanic black women also have a significantly earlier age at diagnosis, high grade tumors, and a higher proportion of triple-negative breast cancers (Chu et al., 2003; Newman, 2005). The focus on TNBC is crucial as there is a lack of therapeutic options for this specific typology. Thus, if one group is disproportionately affected, the results can be devastating. No studies have examined racial disparities across the US with the USCS database, as few have looked beyond the scope of one state.

The research questions of interest are: Are the underlying distributions of age, race and stage at diagnosis different for women with TNBC compared to women with all other types of breast cancer? Do the odds of TNBC diagnosis among women with breast cancer differ by race, age or stage at diagnosis, at the individual level? The research hypotheses are as follows: Non-Hispanic black women will have higher odds of TNBC diagnosis than their non-Hispanic white counterparts in nationally aggregated data analysis. Younger women will have higher odds of TNBC diagnosis and those diagnosed at late stage will have higher odds of TNBC diagnosis.

A descriptive epidemiologic study of the population dataset will be conducted to determine whether there are non-Hispanic black-non-Hispanic white disparities between cases with BC, and subcases with TNBC diagnoses. The unit of analysis is the individual. Proportions and confidence intervals will be calculated and tested, using Chi-Square tests of independence to detect differences in distributions of age groups and race. T tests will be used to detect differences in average age. Bivariate and multivariate logistic regression analyses will be performed to calculate odds ratios and confidence intervals comparing non-Hispanic black and
non-Hispanic white differences, age group differences, and stage differences in diagnosis across states of triple-negative diagnosis among breast cancer cases. This approach is similar to the Bauer and colleagues (2007) California registry study.

**Study 2. Examination of Geospatial Patterns in Triple-Negative Breast Cancer and Factors across the United States**

Geographic disparities in breast cancer and its late-stage diagnosis have been established in the literature for several years (Henley, King, German, Richardson, & Plescia, 2010; Kerner, Andrews, Zauber, & Struening, 1988; Siegel et al., 2018) Kerner and colleagues (1988) called for the use of spatial analysis to inform prevention strategy and policy implementation. Since then, a few studies have utilized spatial analysis in the evaluation of breast cancer outcomes overall, and none have evaluated the TNBC typology (Sheehan et al., 2004; Wang, McLafferty, Escamilla, & Luo, 2008; Wang, Burau, Fang, Wang, & Du, 2008). Additionally, no such studies focused on the United States have utilized Anselin’s Local Moran’s I (LISA) statistics in the analysis (Anselin, 1995). While other methods look for epicenters in a global pattern, the LISA methodology accounts for local spatial instabilities in overall patterns of global spatial association. This methodology is considered more reliable for inference in both the absence and presence of spatial autocorrelation, allowing for the identification of concentrations of both high and low values, and spatial outliers. Examining the spatial distribution of this aggressive triple-negative subtype at the county level provides insight at policy-actionable geographic levels. Use of the comprehensive USCS database provides population-based results that are completely generalizable. The proposed study seeks to fill several methodological and data-based gaps in the literature.
The research questions are as follows: Is there evidence of geospatial disparities in both breast cancer and triple negative breast cancer diagnosis across the United States? Do these clusters coincide? Can residential segregation and socioeconomic conditions predict triple-negative diagnosis? The research hypotheses are: There will be evidence of observed spatial patterns (clustering) of higher-than-average and lower-than-average TNBC rates across the United States. TNBC county rates will be associated with indicators of community disadvantage. Presuming the presence of spatial clusters will lead to efforts to determine socio-ecological and environmental characteristics associated with observed spatial patterns (e.g. socioeconomic conditions and residential segregation).

Descriptive maps will be generated to demonstrate the distribution of TNBC across the United States. The unit of analysis is the county. Spatial clustering is expected when examining geospatial disparities. The first step is to determine the degree of global clustering using the Global Moran’s I statistic. The null hypothesis states that the attribute (i.e., TNBC rates) is randomly distributed among features (i.e., counties) in the study area – the contiguous United States. The larger the Moran’s I statistic, the greater the local area variation observed in the disease rates. A rejection of the hypothesis of spatial randomness with the Moran’s test predicates use of the Local-Indicators of Spatial Analysis (LISA) test for identification of local clusters. Positive spatial autocorrelation in TNBC rates among counties is represented by both high-high and low-low clusters, while negative spatial autocorrelation is represented by high-low and low-high clusters.

Community characteristics will be represented using data from the Spatial Impact Factor Database. The database is supplemented by additional data from the Area Health Resource File and Pennsylvania State University (Bureau of Health Workforce, 2017; Rupasingha, Goetz, &
RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER

Freshwater, 2006). This portion of the study will examine how one’s social environment may correlate with rates of breast and triple negative breast cancer. Spatial regression techniques will be employed to evaluate the association between county level social environment factors and aggregate triple negative breast cancer rates.

**Study 3. Multi-level Analysis of Person-, County- and State-Level Contributors to Triple-Negative Breast Cancer Diagnosis among Women in the United States**

The use of the mixed models is a novel approach to ecological-type regressions that investigate risk factors (e.g. socioeconomic conditions, residential segregation, social capital, housing adequacy from a health standpoint, access to primary care healthcare providers, and state insurance mandates) related to breast cancer outcomes (Liu, Wall, & Hodges, 2005). This could provide greater insights and explain substantially more of the variation in the observed outcome. Multilevel modeling will be employed to explain the variation in TNBC diagnosis at the individual level using person, county and state level predictors (e.g. socioeconomic conditions, residential segregation, access to primary care healthcare providers, and state insurance mandates). The aggregate level variables will include the community characteristics such as socioeconomic conditions, residential segregation and social capital, and state-level characteristics such as breast cancer screening mandates, and implementation of Medicare/Medicaid expansion or state-specific restrictions on Nurse Practitioner or Physician Assistant scope of practice. With this enhanced knowledge and effective dissemination and translation of it, policies and interventions can be designed and targeted to address the barriers and gaps that contribute to the observed TNBC outcomes. Model building for multilevel analysis will allow us to explore how these predictors contribute to diagnosis of TNBC among BC patients.
The research questions are as follows: What is the predicted probability of triple negative breast cancer diagnosis, given breast cancer diagnosis, among females in an average county in the United States? What community variables are strong predictors of triple negative breast cancer diagnosis among breast cancer cases? When we control for patient, county and state characteristics, what is the relationship between patient race and odds of TNBC diagnosis, given breast cancer diagnosis? The research hypotheses are: TNBC diagnosis will vary for females across counties and counties within states and will be associated with several factors reflecting community disadvantage. We anticipate that the odds of triple negative breast cancer diagnosis will remain higher for Non-Hispanic black women compared to Non-Hispanic white women, even when controlling for county and state level characteristics. Multilevel analysis will explain the variation in diagnosis within and between racial groups using person, county and state level predictors.
Descriptive Analysis of Black/White Disparities in Triple Negative Breast Cancer for the US – A Population Based Study from the USCS database

Lia C Scott¹, MPH, Lee Mobley¹, PhD, Tzy-Mey Kuo², PhD, Dora Il’yasova¹, PhD

¹Georgia State University, School of Public Health, Atlanta, GA
²University of North Carolina, Lineberger Cancer Center, Chapel Hill, NC

Correspondence:
Lia C Scott, MPH
Georgia State University
School of Public Health
PO Box 3995
Atlanta, GA 30303

Lscott19@gsu.edu

Manuscript Word Count: 2344
Manuscript Figure Count: 1
Manuscript Table Count: 4
Abstract

Triple negative breast cancer has been associated with a more aggressive histology, poorer prognosis and nonresponsiveness to hormone therapy. Due to the lack of therapeutic options for this cancer type, it is imperative that cancer research identify factors that drive disparities and focus on prevention. This study expands upon the literature by examining the outcome in population-setting rather than a sample, which can validate previous findings, by capturing the majority of the population. Using the United States Cancer Statistics database, we identified 1,151,724 cases of breast cancer from 2010-2014, with the triple negative phenotype accounting for approximately 8.4% of all cases. The underlying distribution of age, race, and stage were statistically significantly different when we compared triple negative breast cancer cases to all other breast cancer cases. Unadjusted and adjusted logistic regression results found that Non-Hispanic black and Hispanic women had higher odds of diagnosis when compared to non-Hispanic white women, with non-Hispanic black women having over twice the odds of diagnosis. Additionally, those less than 50 years old had higher odds of diagnosis while those over 64 had lower odds, compared to age 50 to 64. Women younger than 40 had the highest odds of diagnosis, as compared to the referent group, with an odds ratio of approximately 1.8. Diagnosis at stage III and beyond conferred higher odds of diagnosis of triple-negative breast cancer. In adjusted analyses, these disparities persisted. A subset analysis was conducted on just non-Hispanic black and non-Hispanic white cases to explore the interaction of age, race and stage. This subset accounted for approximately 86% of the breast cancer population. Adjusted logistic regressions were run with age, race and stage as predictors of triple negative breast cancer diagnosis. Interaction effects of age and stage by race were explored. Stage and race were statistically significant moderators of the relationship between age and diagnoses of triple
negative breast cancer. As age increased the odds of triple negative diagnosis decreased, however those diagnosed at late stage had higher odds of triple negative breast cancer compared to those diagnosed in early stage. Additionally, non-Hispanic black women consistently had twice the probability of triple negative diagnosis. This study shows that there is significant burden of disease in triple negative breast cancer for women of color, specifically non-Hispanic black women, and younger women. Additional studies need to be conducted to determine what may be driving these disparities between race, age and stage.
RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER

Introduction

Triple negative breast cancer (TNBC) accounts for approximately 15% of all breast cancer cases, and is associated with aggressive histology, poorer prognosis, shorter survival, and unresponsiveness to usual hormone therapy (Bauer, Brown, Cress, Parise, & Caggiano, 2007). A study conducted on a sample of 51,074 women using the California cancer registry found that the triple negative phenotype was statistically significantly associated with younger age, African American race/ethnicity, later stage diagnosis, lower SES and shortened survival (Field et al., 2005). This study is one of a few that include personal SES data. Non-Hispanic black women also had a significantly earlier age at diagnosis, high grade tumors, and a higher proportion of triple-negative breast cancers (Chu, Lamar, & Freeman, 2003; Newman, 2005). Chu and colleagues study had a sample size of 107,612. The focus on TNBC is crucial as there is a lack of therapeutic options for this specific typology, thus research must shift focus from treatment to identifying risk factors, albeit environmental or genetic, that exacerbate disparities in breast cancer diagnosis in order to create and implement more efficacious population-based prevention strategies.

Although research has established multiple individual-level risk factors associated with general breast cancer diagnoses, racial disparities still exist, particularly between Non-Hispanic white and Non-Hispanic black populations. While age-adjusted incidence rates are higher in Non-Hispanic white women, mortality rates are higher in Non-Hispanic black women. Age-adjusted incidence rates from 2011-2015 were 128.6 per 100,000 for non-Hispanic white women and 126.9 per 100,000 for non-Hispanic black women. The age-adjusted mortality rates from breast cancer in 2011-2015 for Non-Hispanic black women was 28.7 per 100,000 and 20.3 deaths per 100,000 for Non-Hispanic white women (Noone et al., 2018). Research indicates that
survival in Non-Hispanic black women may be worse due to a higher frequency of adverse histologic features (Bauer et al., 2007). When examining cases from 2008 to 2014, Non-Hispanic white women had a higher percentage of localized disease when compared to Non-Hispanic black women (63% vs 54%), yet the 5-year relative survival rate for Non-Hispanic black women was lower in terms of localized (99.1% vs 95.4%), regional (86.4% vs 76.6%), and distant stages (28.1% vs.19.7%) (Noone et al., 2018).

No studies have examined racial disparities in TNBC across the US with the USCS database, as few have looked beyond the scope of one state. Previous studies have found that age and race confer a higher risk of breast cancer diagnosis (Howlader et al., 2013). Older age and has been linked to diagnosis with breast cancer, but this relationship is reversed when we look further into triple negative breast cancer (Siegel, Miller, & Jemal, 2018). In terms of triple negative diagnosis, this type occurs more frequently in younger women and in non-Hispanic black women (Bauer et al., 2007; Dent et al., 2007; Haffty et al., 2006; Harris et al., 2006; Morris et al., 2007). These findings have been limited due to their small sample sizes, the smallest of which was 474 cases from a clinical trial (Harris et al., 2006) to the largest of 197,274 cases over a 10-year period (Morris et al., 2007), thus they are neither spatially representative or generalizable.

With the use of a population dataset this paper aims to validate previous findings in the literature, confirming the proportion of breast cancer cases that are triple negative, and the effect of age, race and stage on diagnosis. The research questions of interest are: Are the underlying distributions of age, race and stage at diagnosis different for women with TNBC compared to women with all other types of breast cancer? Do the odds of TNBC diagnosis among women
with breast cancer differ by race, age or stage at diagnosis, at the individual level? Is there a difference in odds of diagnosis when we compare late stage to distant stage?

The research hypotheses are as follows: Non-Hispanic black women will have higher odds of TNBC diagnosis than their non-Hispanic white counterparts in nationally aggregated data analysis. Younger women will have higher odds of TNBC diagnosis and those diagnosed at late and distant stage will have higher odds of TNBC diagnosis.

Methods

We examined all breast cancer cases diagnosed during 2010–2014 from the United States Cancer Statistics (USCS) database, which is a population-based surveillance system of cancer registries with data representing 99% of the U.S. population (Richardson, Henley, Miller, Massetti, & Thomas, 2016). Most states participate in the USCS registry data system, but five did not provide county-level breast cancer data – Illinois, Kansas, Michigan, Minnesota and Missouri – and four did not code for triple negative data – Connecticut, Iowa, New Mexico, and Utah. Alaska and Hawaii were excluded from analysis due to missing contextual data.

The dataset was analyzed using SAS Software (SAS 9.4, SAS Institute Inc., Cary, NC). Triple negative cases were identified using site specific factors 1, 2 and 15. Late stage was defined as diagnosis at AJCC Stage III and beyond, while distant stage is defined as diagnosis at AJCC Stage IV. Age groups were defined as less than 40, 40 – 49, 50 - 64, 65 – 74, and 75 and older with age 50 -64 serving as the referent group. There were six race/ethnicity categories in the study: non-Hispanic white, Hispanic, non-Hispanic black, American Indian/Alaskan Native, Asian, and Other, with non-Hispanic white serving as the referent category. Descriptive statistics were calculated for age, race and stage variables in the dataset. Chi-Square tests and the student’s t-tests were employed to compare differences in the distribution of age, race, and stage in triple
negative cases versus all other breast cancer cases. Logistic regression was then used to
determine the odds of diagnosis of triple negative breast cancer given breast cancer and its
variation by race, age and stage. Adjusted models were run, and late stage and distant stage were
included in separate models to reduce confounding effects, as the inclusion of both
simultaneously may diminish the effect of each variable. This approach is similar to the Bauer
and colleagues (2007) who conducted the California registry study. A repeated analysis was
conducted on a subset of cases that only included Non-Hispanic white and Non-Hispanic black
women. The interaction of age and stage, stratified by race, was evaluated for this subset.

Results

We identified 1,151,724 breast cancer cases from 2010-2014, with a mean case age of
61.86. Approximately 75% of the cases were non-Hispanic white women, 27.68% were
diagnosed late stage and approximately 5% were diagnosed at distant stage. In this time period,
triple negative cases accounted for 8.4% of all breast cancer cases (Table 1). Results from Chi-
Squared tests demonstrate that there is a statistically significant relationship between the
distributions of race, age group, late stage, and distant stage and triple negative breast cancer
diagnosis. The triple negative group had a statistically significant lower mean age, 59.3,
compared to the other breast cancer group, 62.1 (Table 2).

Race, age and late stage were all statistically significant predictors in the unadjusted and
adjusted logistic regression model (Table 3). In unadjusted models, compared to non-Hispanic
white women, Non-Hispanic black, Hispanic and American Indian/Alaskan Native women had
statistically significant higher odds of TNBC diagnosis while Asian and Other women had lower
odds. Non-Hispanic black women had the highest odds of diagnosis (OR=2.27 (95% CI =2.23,
2.31)), while other women had the lowest odds of diagnosis (OR=0.71 (95% CI=0.64, 0.77)).
RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER

compared to Non-Hispanic white women. Out of the age groups, compared to women age 50-64, women under the age of 40 had the highest odds of TNBC diagnosis, (OR= 1.95 (95% CI=1.90, 2.01)) while those age 75 and older had the lowest odds of diagnosis of TNBC (OR= 0.75 (95%CI=0.73, 0.76)). Women diagnosed at late stage were 69% more likely to be diagnosed with triple negative breast cancer (OR= 1.69 (95%CI=1.68, 1.72)), and women diagnosed at distant stage were 47% more likely to be diagnosed with TNBC (OR= 1.47 (95%CI=1.43, 1.51)).

In the adjusted model for late stage, only Hispanic women did not have a difference in odds of TNBC diagnosis compared to non-Hispanic white women. Non-Hispanic black women had 2.1 times the odds of diagnosis with triple negative breast cancer. The youngest age group had the highest odds of TNBC diagnosis, while the oldest had the lowest odds. Women age 40-49 did not have statistically significantly different odds of diagnosis compared to women age 50-64. For those diagnosed at late stage, the odds of triple negative diagnosis were 1.58 times the odds for those diagnosed earlier than stage three. In the adjusted model for distant stage, the results were similar. Women age 40-49 had a slightly higher odds of diagnosis with TNBC, 1.09, while the results remained the same for Hispanic women. Those diagnosed at distant stage had 1.39 times the odds of diagnosis of TNBC.

In the subset analysis, race, age and stage were used to predict TNBC diagnosis among non-Hispanic black and non-Hispanic white women only. The subset had a size of 973,293 women. All variables in the adjusted model were statistically significant, including the interaction effect (Table 4). When stratified by race, the predicted probability of diagnosis for TNBC remained higher for non-Hispanic black women than for non-Hispanic white women regardless of stage at diagnosis (Figure 1). When we examine probabilities of diagnosis at late-stage for non-Hispanic black and non-Hispanic white women, non-Hispanic black women
consistently had a higher probability of diagnosis. From age 34.8 (2 SD below the mean) to 88.8 (2 SD above the mean), the probability of diagnosis with TNBC for non-Hispanic black women decreased from 24.4% to 15.4%, while the probability of diagnosis with TNBC for non-Hispanic white women decreased from 12.4% to 8.2%. The effect of age on stage is slightly greater for non-Hispanic black women when we compare coefficients (-0.013 v -0.01). Since the magnitude of the slopes is small, the slope does not appear to drastically change in Figure 1. However, the slope for non-Hispanic black women is twice as large (-0.006) compared to NHW women (-.003). The probability of TNBC diagnosis for Non-Hispanic white women at the mean age (61.8) and late stage is 0.10, while for non-Hispanic black women, it is 0.19. The predicted difference in the log-odds of diagnosis for non-Hispanic black and non-Hispanic white women is 0.374, on average.

Discussion

In the multi-state subset of the breast cancer population, approximately, 8.5% of cases were classified as triple negative using site-specific factors. These results find that triple negative cases account for fewer breast cancer cases (8.5%) than found in previous studies, with smaller samples focused on a single state, that show estimates between 12 and 15% (Bauer et al., 2007; Brewster, Chavez-MacGregor, & Brown, 2014; Gretchen, Burke, & Anderson, 2010; Lund et al., 2008; Parise, Bauer, & Caggiano, 2010; Stark et al., 2010; Trivers et al., 2009). The distribution of race was different for the triple-negative cases compared to all other breast cancer cases (p<.0001). Non-Hispanic black women accounted for 10.9% of the other breast cancer cases, but 21.4% of the triple negative cases, while non-Hispanic white women only accounted for 65.7% of triple negative cases. In Bauer and colleagues study (2007), non-Hispanic black women accounted 4.4% of other breast cancer cases and 10% of triple negative cases.
In the present study, age group distributions were statistically significantly different between other breast cancer and triple negative breast cancer cases (p<.0001). The youngest age group, less than 40, accounted for 3.8% of other breast cancer cases, and 7.7% of triple negative cases. In the most comparable registry study (Bauer, 2007), this age group accounted for more breast (5.7%) and triple negative breast cancer cases (12.2%). The proportion of those diagnosed at late-stage and distant stage was statistically significantly higher in the triple negative group compared to the other breast cancer cases. Late stage diagnosis occurred in 37.9% of triple negative cases, and distant stage diagnosis occurred in 6.6% of triple negative cases. This finding is contrary to the Bauer study that found that late stage cases of triple negative breast cancer account for approximately 15% of the cases and distant stage accounted for 4%. We found evidence of different distributions in age, race and stage at diagnosis compared to previous studies. These stark differences demonstrate the importance of national population-based studies.

Overall, this descriptive analysis confirms disparities previously found in the literature and shows that there are significant burdens among women of color, specifically non-Hispanic black women, younger women, and women diagnosed at a later stage when it comes to triple-negative breast cancer diagnosis. This study found that those burdens are higher among these groups than previously estimated, potentially due to the use of a more comprehensive population. These differences were confirmed in the logistic regression analyses. In both adjusted and unadjusted models, non-Hispanic black women had significantly higher odds of triple negative diagnosis compared to non-Hispanic white women. The youngest age group also had significantly higher odds of triple negative diagnosis. Women diagnosed at late and distant stage had significantly higher odds of triple negative diagnosis. The subset analysis of just non-Hispanic black and non-Hispanic white women shows that disparities persist between these two
groups. In our study, we found that non-Hispanic black women have twice the probability of diagnosis with triple negative breast cancer, when controlling for age and stage at diagnosis. The effect of age and stage at diagnosis differed by race group. Younger age conferred higher odds of diagnosis for both Non-Hispanic black and Non-Hispanic white women, however there is a larger gap between the probabilities of diagnosis by stage status for non-Hispanic black women.

Given the large sample size and geospatial coverage of the data, these results are somewhat different and also more generalizable than previous studies. The database needs to be monitored for a shift in policy and data quality that would allow for the addition of states not currently included in the analysis. Considering these results, it is important to consider what additional factors may influence individual level variations in diagnosis. Further exploration into additional individual and environmental characteristics is necessary to identify what may be driving these disparities in diagnosis. Due to the aggressive nature of triple-negative breast cancer, and lack of therapeutic options, it is important to know which groups confer a higher risk to better provide intervention.
Tables

Table 1. Descriptive Statistics of Study Population, n=1151724

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NHW</td>
<td>862205</td>
<td>74.86</td>
</tr>
<tr>
<td>Hispanic</td>
<td>95507</td>
<td>8.29</td>
</tr>
<tr>
<td>NHB</td>
<td>135389</td>
<td>11.76</td>
</tr>
<tr>
<td>AI/AN</td>
<td>5504</td>
<td>0.48</td>
</tr>
<tr>
<td>Asian</td>
<td>44424</td>
<td>3.86</td>
</tr>
<tr>
<td>Other</td>
<td>8695</td>
<td>0.75</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age Groups, mean</td>
<td>1151724</td>
<td>61.84</td>
</tr>
<tr>
<td>&lt;40</td>
<td>47329</td>
<td>4.11</td>
</tr>
<tr>
<td>40-49</td>
<td>179715</td>
<td>15.6</td>
</tr>
<tr>
<td>50-64</td>
<td>433798</td>
<td>37.67</td>
</tr>
<tr>
<td>65-74</td>
<td>275982</td>
<td>23.96</td>
</tr>
<tr>
<td>75+</td>
<td>214900</td>
<td>18.66</td>
</tr>
<tr>
<td><strong>Late Stage</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>310588</td>
<td>27.68</td>
</tr>
<tr>
<td>No</td>
<td>811633</td>
<td>72.32</td>
</tr>
<tr>
<td><strong>Distant</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>54073</td>
<td>4.82</td>
</tr>
<tr>
<td>No</td>
<td>1068148</td>
<td>95.18</td>
</tr>
<tr>
<td><em>Missing</em></td>
<td>29503</td>
<td>2.6</td>
</tr>
<tr>
<td><strong>TNBC</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>96749</td>
<td>8.4</td>
</tr>
<tr>
<td>No</td>
<td>1054975</td>
<td>91.6</td>
</tr>
</tbody>
</table>
### Table 2. Differences in Frequencies of Age, Race, and Stage for Other Breast Cancers and Triple Negative Breast Cancer

<table>
<thead>
<tr>
<th>Variable</th>
<th>Other BC n=1054975</th>
<th>TNBC n=96749</th>
<th>Test for Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Percent</td>
<td>N</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NHW</td>
<td>798626</td>
<td>75.70</td>
<td>63579</td>
</tr>
<tr>
<td>Hispanic</td>
<td>87069</td>
<td>8.25</td>
<td>8438</td>
</tr>
<tr>
<td>NHB</td>
<td>114663</td>
<td>10.87</td>
<td>20726</td>
</tr>
<tr>
<td>AI/AN</td>
<td>5002</td>
<td>0.47</td>
<td>502</td>
</tr>
<tr>
<td>Asian</td>
<td>41382</td>
<td>3.92</td>
<td>3042</td>
</tr>
<tr>
<td>Other</td>
<td>8233</td>
<td>0.78</td>
<td>462</td>
</tr>
<tr>
<td>Age Groups</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;40</td>
<td>39855</td>
<td>3.78</td>
<td>7474</td>
</tr>
<tr>
<td>40-49</td>
<td>162650</td>
<td>15.42</td>
<td>17065</td>
</tr>
<tr>
<td>50-64</td>
<td>395805</td>
<td>37.52</td>
<td>37993</td>
</tr>
<tr>
<td>65-74</td>
<td>256176</td>
<td>24.28</td>
<td>19806</td>
</tr>
<tr>
<td>75+</td>
<td>200489</td>
<td>19.00</td>
<td>14411</td>
</tr>
<tr>
<td>Mean Age*</td>
<td>62.07</td>
<td>59.28</td>
<td>t=59.86</td>
</tr>
<tr>
<td>Late Stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>273929</td>
<td>25.97</td>
<td>36659</td>
</tr>
<tr>
<td>No</td>
<td>752341</td>
<td>71.31</td>
<td>59292</td>
</tr>
<tr>
<td>Distant</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>47661</td>
<td>4.52</td>
<td>6412</td>
</tr>
<tr>
<td>No</td>
<td>978609</td>
<td>92.76</td>
<td>89539</td>
</tr>
<tr>
<td>Missing**</td>
<td>28705</td>
<td>2.72</td>
<td>798</td>
</tr>
</tbody>
</table>

* Satterthwaite T-test used to compare mean age differences ($F=1.08$, $p<.0001$ test for equal variance). Percent is the mean age for each group.

** This represents the number of cases that were missing stage data.
Table 3. Association between TNBC diagnosis and Race/Ethnicity, Age and Stage at Diagnosis

<table>
<thead>
<tr>
<th>Variable</th>
<th>Odds Ratio</th>
<th>95% C.I.</th>
<th>p value*</th>
<th>Odds Ratio</th>
<th>95% C.I.</th>
<th>p value*</th>
<th>Odds Ratio</th>
<th>95% C.I.</th>
<th>p value*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unadjusted</td>
<td></td>
<td></td>
<td>Adjusted - Late Stage</td>
<td></td>
<td></td>
<td>Adjusted - Distant Stage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NHW</td>
<td>REF</td>
<td></td>
<td></td>
<td>REF</td>
<td></td>
<td></td>
<td>REF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td></td>
<td></td>
<td></td>
<td>1.217</td>
<td>1.189, 1.247</td>
<td>0.001</td>
<td>1.106</td>
<td>1.08, 1.134</td>
<td>0.3624</td>
</tr>
<tr>
<td>NHB</td>
<td></td>
<td></td>
<td></td>
<td>2.271</td>
<td>2.233, 2.309</td>
<td>&lt;.0001</td>
<td>2.111</td>
<td>2.075, 2.148</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>AI/AN</td>
<td></td>
<td></td>
<td></td>
<td>0.923</td>
<td>0.889, 0.959</td>
<td>&lt;.0001</td>
<td>0.851</td>
<td>0.819, 0.885</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Asian</td>
<td></td>
<td></td>
<td></td>
<td>0.705</td>
<td>0.642, 0.774</td>
<td>&lt;.0001</td>
<td>0.714</td>
<td>0.646, 0.788</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td>1.261</td>
<td>1.15, 1.382</td>
<td>0.0174</td>
<td>1.188</td>
<td>1.083, 1.305</td>
<td>0.035</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.923</td>
<td>0.889, 0.959</td>
<td>&lt;.0001</td>
<td>0.851</td>
<td>0.819, 0.885</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.705</td>
<td>0.642, 0.774</td>
<td>&lt;.0001</td>
<td>0.714</td>
<td>0.646, 0.788</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Age Groups</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;40</td>
<td></td>
<td></td>
<td></td>
<td>1.954</td>
<td>1.902, 2.007</td>
<td>&lt;.0001</td>
<td>1.759</td>
<td>1.711, 1.808</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>40-49</td>
<td></td>
<td></td>
<td></td>
<td>1.093</td>
<td>1.072, 1.114</td>
<td>&lt;.0001</td>
<td>1.066</td>
<td>1.046, 1.087</td>
<td>0.0183</td>
</tr>
<tr>
<td>50-64</td>
<td></td>
<td></td>
<td></td>
<td>0.374</td>
<td>0.343, 0.406</td>
<td>&lt;.0001</td>
<td>0.348</td>
<td>0.323, 0.374</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>65-74</td>
<td></td>
<td></td>
<td></td>
<td>0.805</td>
<td>0.791, 0.82</td>
<td>&lt;.0001</td>
<td>0.831</td>
<td>0.810, 0.852</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>75+</td>
<td></td>
<td></td>
<td></td>
<td>0.749</td>
<td>0.734, 0.764</td>
<td>&lt;.0001</td>
<td>0.739</td>
<td>0.720, 0.759</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Late Stage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td>1.698</td>
<td>1.675, 1.722</td>
<td>&lt;.0001</td>
<td>1.58</td>
<td>1.558, 1.602</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td>REF</td>
<td></td>
<td></td>
<td>REF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td>1.47</td>
<td>1.431, 1.511</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
<td>1.388</td>
</tr>
<tr>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td>REF</td>
<td></td>
<td></td>
<td>REF</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Adjusted Logistic Regression Results for Non-Hispanic black/Non-Hispanic white Subset Analysis, n=973293

<table>
<thead>
<tr>
<th>Variable</th>
<th>Est.</th>
<th>S.E.</th>
<th>p value*</th>
<th>Est.</th>
<th>S.E.</th>
<th>p value*</th>
<th>Est.</th>
<th>S.E.</th>
<th>p value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-1.365</td>
<td>0.017</td>
<td>&lt;.0001</td>
<td>-1.785</td>
<td>&lt;.0001</td>
<td>-0.845</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NHW</td>
<td>REF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NHB</td>
<td>0.374</td>
<td>0.004</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>-0.011</td>
<td>0.000</td>
<td>&lt;.0001</td>
<td>-0.010</td>
<td>0.000</td>
<td>&lt;.0001</td>
<td>-0.013</td>
<td>0.007</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Stage</td>
<td>0.136</td>
<td>0.017</td>
<td>&lt;.0001</td>
<td>0.139</td>
<td>0.020</td>
<td>&lt;.0001</td>
<td>0.085</td>
<td>1.089</td>
<td>0.013</td>
</tr>
<tr>
<td>Age*Stage</td>
<td>0.002</td>
<td>0.000</td>
<td>&lt;.0001</td>
<td>0.002</td>
<td>0.000</td>
<td>&lt;.0001</td>
<td>0.003</td>
<td>1.003</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>
Figure 1. Interaction Effects of Stage at Diagnosis on Age at Diagnosis by Race
RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER

References


44
RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER


RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER


Chapter 3
Research Study 2

Manuscript 2

Examination of Geospatial Patterns in Triple-Negative Breast Cancer and Factors across the United States

Lia C Scott¹, MPH, Lee Mobley¹, PhD, Tzy-Mey Kuo², PhD, Dora Il’yasova¹, PhD

¹Georgia State University, School of Public Health, Atlanta, GA
²University of North Carolina, Lineberger Cancer Center, Chapel Hill, NC

Correspondence:
Lia C Scott, MPH
Georgia State University
School of Public Health
PO Box 3995
Atlanta, GA 30303

Lscott19@gsu.edu

Manuscript Word Count: 3128
Manuscript Figure Count: 6
Manuscript Table Count: 4
Abstract

The literature amply demonstrates that disparities exist between age, race and stage of diagnosis in triple negative breast cancers, however few studies have addressed how the social environment impacts diagnosis. Evaluation of geospatial patterns reflecting disparities is a crucial part of cancer research. This ecological study evaluates geographic disparities using 1,151,724 cases of breast cancer from 2010-2014 in 2430 counties across 39 states. The data are aggregated to the county-level using FIPS codes recorded at the time of diagnosis. County-level spatial cluster analysis, which is descriptive and spatial regression, which is confirmatory in ecological modeling at the county level, were employed to analyze geographic patterns of crude breast cancer and triple negative breast cancer rates and their contextual predictors. County level contextual factors were drawn from several public-use sources.

Crude rates were defined as the total number of cases from 2010-2014 per 100,000 women. Evidence of spatial autocorrelation in both breast and triple negative breast cancer rates was found and multiple clusters coincided with one another. The geospatial patterns of crude breast cancer and triple negative breast cancer rates were distinct from one another, with most high rate clusters for the former in the North and Northeastern United States, and in the South, for the latter. We identified where the two types of clusters converged, and 159 counties had both high breast and triple negative breast cancer rates, while 97 counties had low breast cancer rates but high triple negative breast cancer rates. In the final spatial error regression model for breast cancer rates, all residential segregation and socioeconomic variables were statistically significant predictors. In the final spatial lag + error model for triple negative breast cancer, only the isolation indices and the diversity index were statistically significant. Both models found that residential segregation played a role in the outcomes. Future studies need to account for spatial
dependence when looking at geographic data, such as that provided by the USCS database, and further explore additional contextual and compositional variables that may contribute to the variation in county-level breast cancer rates.
Introduction

Geographic disparities in the diagnosis of breast cancer has been established in the literature for several years (Carvalho, Bacchi, Pincerato, Van de Rijn, & Bacchi, 2014; Henley, King, German, Richardson, & Plescia, 2010; Kerner, Andrews, Zauber, & Struening, 1988; Mobley, Kuo, Scott, Rutherford, & Bose, 2017; Mobley & Kuo, 2015; Scott, Mobley, & Il’yasova, 2017; Siegel, Miller, & Jemal, 2018; Tian, Wilson, & Zhan, 2011). Kerner and colleagues (1988) called for the use of spatial analysis to inform prevention strategy and policy implementation. Since then, a few studies have utilized spatial analysis in the evaluation of breast cancer (BC) outcomes overall, and none have evaluated the triple-negative subtype (TNBC) (Kuo, Mobley, & Anselin, 2011; Mobley, Kuo, Watson, & Gordon Brown, 2012; Sheehan et al., 2004; Wang, McLafferty, Escamilla, & Luo, 2008). Additionally, no such studies focused on the United States have utilized Anselin’s Local Moran’s I (LISA) statistics in the analysis (Anselin, 1995). While other methods look for epicenters in a global pattern, the LISA methodology accounts for local spatial instabilities in overall patterns of global spatial association. This methodology is considered more reliable for inference in both the absence and presence of spatial autocorrelation, allowing for the identification of concentrations of both high and low values, and spatial outliers. Examining the spatial distribution of this aggressive TNBC typology at the county level provides insight at policy-actionable geographic levels.

Besides comparing the geospatial distribution, it is important to examine underlying community characteristics. This information can provide insight on the mechanisms by which TNBC may be occurring more in certain communities compared to others. Previous studies have compared the distribution of community characteristics between high and low cluster centers, which is basically a descriptive approach (Mobley, Finkelstein, Khavjou, & Will, 2004; Scott et
This study aims to delve further by predicting breast cancer and triple negative breast cancer rates through spatial regression techniques covering the entire geography of the sample, thereby understanding which variables may have the largest effect on these county level rates.

This ecological study focuses on the association of residential segregation, education and income inequality on triple-negative diagnosis rates in counties. Residential segregation is represented by one measure on evenness, the diversity index, and two measures of exposure, the isolation indices. The isolation indices used in the study focused on the distribution of the non-Hispanic black and non-Hispanic white population (Charles, 2003; Massey & Denton, 1988). Education data was drawn from the United States Department of Agriculture Rural Atlas which derives estimates from the United States Census Bureau’s American Community Survey. This variable represented the percent of the population with a four-year college degree or higher. For income inequality, the Gini index of income inequality and the poverty rate were used. The Gini index represents the income distribution of a county’s residents and is one of the widely used measures in the literature on income inequality and health (Chen & Crawford, 2012). A higher Gini index implies a higher level of area income inequality (Fan, Wen, & Kowaleski-Jones, 2016). It is important to include the various dimensions spatial socioeconomic polarization and segregation to manage different aspects that might be confounding if omitted, as few studies have explored the joint impact of spatial economic and racial polarization on population health (Feldman, Waterman, Coull, & Krieger, 2015; Krieger, Singh, & Waterman, 2016). This study is driven by ecosocial theory of disease distribution and the hypothesis that women with social and economic privilege are most likely to have ER+ tumors (Krieger, Singh, & Waterman, 2016).

The research questions are as follows: Is there evidence of geospatial disparities in both breast cancer and triple negative breast cancer diagnosis across the United States? Do these
clusters coincide? Can residential segregation and socioeconomic conditions predict triple-negative diagnosis? The research hypotheses are: There will be evidence of observed spatial patterns (clustering) of higher-than-average and lower-than-average TNBC rates across the United States. TNBC county rates will be associated with indicators of community disadvantage.

**Methods**

We examined all breast cancer cases diagnosed during 2010–2014 from the United States Cancer Statistics (USCS) database, which is a population-based surveillance system of cancer registries with data representing 99% of the U.S. population (Richardson et al, 2016). Most states participate in the USCS registry data system, but five did not provide county-level breast cancer data – Illinois, Kansas, Michigan, Minnesota and Missouri – and four did not code for triple negative data – Connecticut, Iowa, New Mexico, and Utah. Alaska and Hawaii were excluded from analysis due to missing contextual data. Community characteristics will be represented using data from the Spatial Impact Factor Database (Mobley, 2015). The database is supplemented by additional data from the United States Department of Agriculture and Pennsylvania State University (Bureau of Health Workforce, 2017; Rupasingha, Goetz, & Freshwater, 2006).

Independent variables included in the study were non-Hispanic black and non-Hispanic white isolation indices, Theil’s diversity index, the Gini index of income disparity, poverty, and educational attainment. Non-Hispanic black and non-Hispanic white isolation indices are measures of exposure that indicate the probability the specified race group will encounter another person from their race group within the areal unit (Massey & Denton, 1988). A value of 1 indicates a perfectly racially isolated county for the respective race. Theil’s Diversity index is a multi-group measure that reflects the level of diversity within the county. This county level
variable is based on tract-level racial composition. Counties where tracts all have the same composition (all individuals in a population are associated with the same racial group) have a low diversity index while counties where different races or ethnicities are separated into cultural enclaves among the tracts will have high indices (White, 1986). A value of 0 for the diversity index indicates no diversity in the population while a value of 1 indicates maximum diversity, where individuals are evenly distributed among two or more mutually exclusive groups (Roberto, 2015).

The Gini coefficient is the most commonly used indicator of income inequality, and its use here allows for comparisons with other income inequality studies (Farley, 2006; Jones-Smith, Gordon-Larsen, Siddiqi, & Popkin, 2011). The Gini index reflects the degree to which income is equally shared within a county and is one of the widely used measures in the literature on income inequality and health (Chen & Crawford, 2012). An index value of 0 represents a perfectly equal distribution of income while an index value of 1 represents a perfectly unequal distribution of income. Poverty was represented as the proportion of the population in poverty as defined by the U.S. Census. Educational attainment was defined as the percent of the population age 25 and older with a graduate or professional degree (Evenden, Harper, Brailsford, & Harindra, 2006).

The dataset was combined using the county FIPS code and summary statistics were computed using SAS Software (SAS 9.4, SAS Institute Inc., Cary, NC). Triple negative cases were identified using site specific factors 1, 2 and 15. Crude BC rates were defined as the total number of BC cases from 2010-2014 per 100,000 persons, and crude TNBC rates were defined as the total number of TNBC cases from 2010-2014 per 100,000 persons. A shapefile was created and following the approach in Mobley et al (2004) and Schieb et al (2013), the Global
RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER

(Moran’s I) and Local Indicators of Spatial Association (LISA) spatial clustering tests were performed using GeoDa software (Anselin, Bera, Florax, & Yoon, 1996; Anselin, Syabri, & Kho, 2006). The Queen Contiguity matrix was used for spatial weights. This type of matrix defines neighbors as counties that share borders or vertices. Results were mapped in QGIS (free, open source GIS software). The global Moran’s I test determines if there is global clustering in the pattern of TNBC rates but cannot identify the location of the clusters, thus we use the LISA test to identify the local clusters.

The LISA statistic is computed using conditional permutation, or bootstrapping, that holds the value fixed for the county of interest and randomly permutes the remaining values to obtain a reference distribution. The observed value is then compared to the distribution to determine extremity of the value. Thus, LISA statistics are relative to the observations of the variable of interest and clusters are determined significant at the \( p < 0.05 \) level. This was repeated for each county to 99999 permutations and counties are then classified as either non-significant or as falling into four categories, relative to the mean (Anselin, 1995). There are four types of spatial clusters identified using the LISA statistic: high-high (higher than average rates adjacent to higher than average rates), low-low, high-low, low-high. Positive spatial autocorrelation (cluster) in BC and TNBC rates among counties is represented by both high-high and low-low clusters, while negative spatial autocorrelation (outlier) is represented by high-low and low-high clusters (Anselin et al., 1995). Clusters are identified when the observation is more similar to its neighbors as summarized by the spatial lag (weighted average of the neighboring values) (Anselin, 2004). For spatial clusters, the cores and neighbors are of interest, while for spatial outliers, the cores are the actual locations of interest (Anselin, 2004).
RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER

Descriptive maps were created to assess the distribution of BC and TNBC rates across the country (Figures 1 and 2). Additionally, the results of the LISA analyses were mapped. These maps show the center of the cluster in color (i.e., red for high-high), while the actual extent of the cluster includes the center and its surrounding neighbors as defined by the queen weights matrix (Figure 3 and 4). The neighbors are properly included in the cluster, shown here as a grey buffer zone around the center. Co-location maps are presented, that demonstrates where BC and TNBC clusters and centers occur concurrently and diverge (Figures 5 and 6). Descriptive statistics were calculated for the outcome variable, BC and TNBC incidence rates, and the contextual variables of interest (Table 1).

Due to evidence of strong spatial autocorrelation in the spatial cluster analyses, spatial regression techniques were employed to estimate the association with county level contextual factors and both breast cancer rates and triple negative breast cancer rates from 2010-2014 (Tables 3 and 4). Independent variables were assessed for multicollinearity prior to inclusion in the models. In Geoda, we ran a classic spatial regression – a replication of ordinary least squares (OLS) regression with spatial diagnostics, specifically, evaluating the Moran’s I statistic and Lagrange Multipliers (LM) for both lag and error utilizing a queen contiguity matrix for both breast and triple negative breast cancer rates. OLS must meet certain assumptions: the random error terms follow a normal distribution, are independent and have a constant variance. Given the Moran’s I results, the assumption of independence is violated by spatial dependence in the data. The model specification for OLS (in matrix notation) is as follows:

\[ y = X\beta + \varepsilon \quad \varepsilon \sim N(0, \sigma^2 I) \]
where $y$ is the dependent variable with $N$ rows, $X$ is a matrix with observation on $K$ independent variables, $\beta$ is a vector with $K$ regression coefficients, $\varepsilon$ is the random error term, $\sigma^2$ is the error variance and $I$ is an identity matrix of dimension $N$ by $N$ (Chasco, 2013).

In both models, spatial diagnostics show that the Moran’s I of the residuals were statistically significant, the errors were not normally distributed and were heteroscedastic. The error model was appropriate for the breast cancer model. For the spatial error model, we utilized the General Method of Moments (robust to non-normality) estimation with Kelejian & Prucha (KP) Heteroscedastic-consistent errors. This method allows for robust estimation with endogenous explanatory variables in the presence of both spatial heteroscedasticity and autocorrelation (Table 3) (Chasco, 2013; Kelejian & Prucha, 2010). It outputs the coefficient lambda ($\lambda$), which is interpreted as a nuisance parameter. The model specifications with a spatial autoregressive error term is:

$$ y = X\beta + u \text{ for } u = \lambda Wu + \varepsilon $$

where the additional parameters beyond OLS are defined as $Wu$ is the spatial lag of the errors, $\lambda$ is the nuisance parameter and $\varepsilon$ is the error with a mean of 0 and variance matrix $\sigma^2$ (Chasco, 2013).

The lag process was appropriate for the triple negative breast cancer model. For the spatial lag model, we utilized the General Method of Moments estimation with White variance to accommodate the heteroscedasticity (Chasco, 2013; White, 1980). This model outputs the coefficient rho ($\rho$), which is interpreted as a measure of the extent of spatial spillovers. The spatial lag model specification with the spatial lagged dependent variable is:

$$ y = \rho Wy + X\beta + \varepsilon $$
where \( Wy \) as a \( N \) by 1 vector of spatial lags for the dependent variable, \( \rho \) is the spatial autoregressive coefficient and \( \varepsilon \) is a \( N \) by 1 vector of normally distributed random error terms, with mean 0 and homoscedastic variances \( \sigma^2 \) (Chasco, 2013). The Anselin-Kelejian test for spatial dependence was statistically significant for the lag model on TNBC rates, indicating evidence of spatial autocorrelation that was still unaccounted for (Anselin & Kelejian, 1997; Anselin & Rey, 1991). We then employed a combined spatial lag and error model using the General Method of Moments estimation with KP Heteroscedastic errors (Table 4). The model specification for the joint spatial lag and spatial error model is as follow:

\[
y = \rho Wy + X\beta + u \quad \text{for} \quad u = \lambda Wu + \varepsilon
\]

where the additional parameters beyond OLS are defined as \( Wy \) as a \( N \) by 1 vector of spatial lags for the dependent variable, \( \rho \) and \( \lambda \) are the spatial autoregressive coefficients, and \( \varepsilon \) is the error term (Chasco, 2013). The inclusion of \( Wy \) allows us to assess the significance of the explanatory variables while controlling for spatial dependence.

Results

From the participating registries in the USCS database, there were 2430 counties in 39 states included in the analysis. The average county-level breast cancer rate was 86.1 cases per 100,000 (SD = 20.9) while the average county-level triple negative breast cancer rates was 7.7 cases per 100,000 (SD = 4.2). There was a large range of values for each of the outcomes, as demonstrated by Figures 1 and 2. Breast cancer incidence rates ranged from 7.1 to 241.0 cases per 100,000 while triple negative breast cancer incidence rates ranged from 0 to 30.8 cases per 100,000. The Moran’s I statistic for both BC (\( I = 0.29 \)) and TNBC (\( I = 0.22 \)) rates was positive and statistically significant indicating evidence of clustering or positive spatial autocorrelation.
There were 231 positive and 264 negative spatial cluster centers found for breast cancer rates. There were 107 spatial outliers (Figure 3). There were 175 positive and 226 negative spatial cluster centers identified for triple negative breast cancer rates. There were 117 spatial outliers (Figure 4). There were 28 cluster centers that coincided for both high BC and high TNBC rates and 104 cluster centers for both low BC and TNBC rates. There were 5 centers each that diverged, five that were low BC and high TNBC, and five that were high BC and low TNBC (Figure 5). When we include the entire clusters, rather than just the centers, there were 159 counties that were High BC-High TNBC, 324 counties that were Low BC-Low TNBC, 97 counties that were Low BC-High TNBC and 54 counties that were High BC- Low TNBC (Figure 6).

All contextual variables were statistically significant predictors of breast cancer rates in the breast cancer spatial error model. The nuisance parameter, lambda=0.45, reflects the effects of omitted county variables that are similar across the county observations. In the spatial lag model for triple negative rates, only the isolation variables were statistically significant predictors. The spatial autoregressive parameter, rho=0.59, allows us to assess the significance of the independent variables after spatial dependence is controlled for. This parameter reflects the effects of a weighted sum of the values of the dependent variable at other locations (Anselin & Rey, 1991). The significance of the Anselin-Kelejian test indicates that the spatial autocorrelation effect is not completely captured by this model. Both spatial parameters were statistically significant in the final lag + error model. In this model, all residential segregation variables were statistically significant predictors.

Discussion
This study found distinct geospatial patterns in crude breast cancer and triple negative breast cancer incidence rates at the county level. High breast cancer rates were found mainly in the northern United States (US), while high triple negative breast cancer rates were found more in the southern US. This may be related to the geographic distribution of race across the United States. The South and Southeast has the largest concentration of Non-Hispanic black persons among all regions in the United States. Given the higher odds of diagnosis among non-Hispanic black women, we could expect a spike in crude rates by region, due to distinct demographic differences by region. These geospatial patterns were confirmed by the LISA cluster analysis.

The vast majority of High-High clusters for breast cancer rates were found along the Northern region of the eastern seaboard, as well as across the entire northern United States, with a cluster in Northeast California and in Central Florida. The majority of New England and Virginia are a part of High-High clusters. On the contrary, Texas, Nevada and much of the Southeast and South are a part of the Low-Low clusters for this outcome. Out of 2430 counties, 1434 (59.0%) were included in a spatial breast cancer cluster.

The LISA Cluster map for crude triple negative breast cancer rates showed a distinctly different geographic pattern. Results were similar for states like Nevada and Texas, but the Eastern part of the United States essentially flips in pattern. Almost no counties from New England are located within a cluster, with the exception of a small cluster in eastern New York. Virginia, North Carolina, South Carolina, Georgia, Alabama, and Mississippi comprised most of the High-High clusters, with a few clusters located in the Dakotas, Idaho, and Montana. There were less counties in a spatial cluster for triple negative breast cancer rates, n=1225 (50.4%). It is important that future studies explore the dynamics behind spatial outliers, particularly ‘Low-
RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER

High’ counties, to determine mechanisms as to why those counties are faring better than their neighbors.

Of note, there are 97 counties that have Low BC rates and High TNBC rates (Figure 6). These clusters were located on the border of Idaho and Montana, Arkansas, Louisiana, Mississippi, Alabama, Georgia, Florida, Virginia, West Virginia, Nebraska and North Dakota. As evident by Figures 5 and 6, it is imperative to consider the extent of the full cluster rather than just cluster centers. We potentially miss important information by solely focusing on the cluster centers. Future studies should examine predictors of county status and evaluate whether there or demographic, social, or physical environment differences that distinguish counties in the cluster category from other counties.

In this study, we examined the associations of social environment factors with breast and triple negative breast cancer rates. Spatial regression was performed to help assess which factors had the largest effect estimates. Two income related variables, the Gini Index and poverty rate, had the largest associations with the breast cancer rate. We would expect a perfectly unequal county, in regard to the Gini index, to have a 75.6 more BC cases per 100,000 compared to a perfectly equal county, on average. Every 10% increase in poverty corresponds to 6.9 cases per 100,000 decrease in the breast cancer rate, on average. The residential segregation measures for isolation had positive associations while the diversity index, a measure of spatial evenness, had a negative association on the breast cancer rate. The diversity index measures how evenly racial groups are distributed across the county, regardless of the size of each group (Iceland, 2004). The higher the value, the more diverse the county. Compared to a perfectly isolated county (where all tracts have the same composition), we would expect a county with maximum diversity to have 28.1 less BC cases per 100,000 on average. A perfectly segregated non-Hispanic black county is
associated with a 31.8 increase in BC cases per 100,000 compared to a perfectly integrated county, while a perfectly segregated non-Hispanic white county is associated with a 47.5 increase in BC cases per 100,000, on average. Non-Hispanic white isolation has a larger effect on breast cancer rates than non-Hispanic black isolation, but integration (diversity index) considering all racial units corresponds with decreased breast cancer rates. Finally, county-level higher education had a substantially small effect on the outcome. For every 1% increase in the population that has at least a college education, we expect the breast cancer rate to increase by 0.15, on average.

Several of these relationships changed for triple negative breast cancer rates. The isolation indices and diversity index were the largest and only significant predictors of triple negative breast cancer rates and income-related variables were no longer significant. A perfectly segregated non-Hispanic Black county is associated with a 4.6 increase in TNBC cases per 100,000 compared to a perfectly integrated county, while a perfectly segregated non-Hispanic White county is associated with a 2.2 increase in TNBC cases per 100,000, on average. The diversity index was negatively associated with the outcome. A perfectly diverse county, with multiple racial groups distributed among tracts within the county, is expected to have 2.8 less TNBC cases per 100,000 than a perfectly isolated county, on average. The average influence of county observations by their neighboring observations is represented by rho = 0.634, while the nuisance parameter = -0.539.

When using spatial data, it is important to consider spatial dependencies in analysis. Observations are correlated with others that are spatially proximate, violating the assumption of independence and resulting in spill-over of information. Failing to account for spatial dependencies may obscure effects that are indeed present given the data and can impart
misspecification bias. In this case, spatial diagnostics predicate the use of a spatial error model for breast cancer rates and a spatial lag + error model for triple-negative breast cancer rates. In the breast cancer model, if spatial autocorrelation is ignored, the estimators may be unbiased yet inefficient. Consequentially, inference based on the variance would be biased (Anselin & Rey, 1991). The spatial error model is the correct specification when neighbors react similarly due to a common underlying phenomenon. The spatial lag or autoregressive model is the correct specification for a diffusion phenomenon among activities of adjacent neighbors. In the triple negative model, if spatial autocorrelation is ignored, then the estimators may be biased, along with the misspecification issues drawn from a spatial error process.

Consequently, given the ecological nature of the study, we cannot infer individual level behavior from the aggregate rates. Inclusion of individual level variables along with county level variables may allow us to account for more variation in the data. Additionally, the use of crude rates limits the interpretation of this analysis as crude rates may be a poor measure of underlying risk and can spuriously identify outliers in counties with very small populations (Anselin, Sridharan, & Gholston, 2007). The literature has also demonstrated that TNBC is associated with age. Appropriate population denominators were not available in the dataset to create age-specific rates in order to directly adjust the crude rates.

Overall, this study demonstrates that spatial dependencies exist when we examine both outcomes at an aggregate geographic level. To our knowledge, no previous studies have examined geospatial clustering of breast and triple negative breast cancer, their co-locations, or conducted spatial regression to evaluate context. It is imperative that we continue to explore what is causing these spatial processes to better develop prevention and intervention methods. Given the ecological nature of the study, it is important to make inferences based around context.
RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER

Characteristics of the counties that had both low breast and triple negative breast cancer rate could inform policy as to what is working to protect the women in these communities. An evaluation of differences between those counties and counties where both outcomes are high, or breast cancer is low and triple negative is high can further elucidate factors that may be driving these geographic disparities.
RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER

Figure 1. Descriptive Map of 2010-2014 overall breast cancer rates defined by quartile breaks

Figure 2. Descriptive Map of 2010-2014 triple-negative breast cancer rates defined by quartile breaks
RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER

Figure 3. LISA Cluster Map of 2010-2014 overall breast cancer rates

Figure 4. LISA Cluster Map of 2010-2014 triple-negative breast cancer rates
RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER

Figure 5. Coincident LISA Clusters Centers 2010-2014 for BC and TNBC rates

Figure 6. Coincident LISA Clusters 2010-2014 for BC and TNBC rates
### Table 1. Descriptive Statistics for County Level Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Variable Description</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>R_crude_BC</td>
<td>Crude breast cancer rate (cases per 100,000 females)</td>
<td>86.0783</td>
<td>20.9117</td>
<td>7.080</td>
<td>240.964</td>
</tr>
<tr>
<td>R_crude_TN</td>
<td>Crude triple negative breast cancer rate (cases per 100,000 females)</td>
<td>7.7351</td>
<td>4.1726</td>
<td>0</td>
<td>30.769</td>
</tr>
<tr>
<td>Iblack10</td>
<td>Isolation Index: probability that non-Hispanic black persons will meet other non-Hispanic black persons within defined area, 2010</td>
<td>0.26251</td>
<td>0.20741</td>
<td>0</td>
<td>0.8647</td>
</tr>
<tr>
<td>Iwhite10</td>
<td>Isolation Index: probability that non-Hispanic white persons will meet other non-Hispanic white persons within defined area, 2010</td>
<td>0.73452</td>
<td>0.15163</td>
<td>0.03154</td>
<td>1</td>
</tr>
<tr>
<td>Divr10</td>
<td>Diversity Index (Theil Index): measures the even-ness or uneven-ness of the spatial distribution of population subgroups in tracts within areas (counties), 2010</td>
<td>0.1981</td>
<td>0.0916</td>
<td>0</td>
<td>0.69685</td>
</tr>
<tr>
<td>G25UP1014</td>
<td>percent of population aged 25+ years with a graduate or professional degree, 2010-2014</td>
<td>29.56426</td>
<td>10.65997</td>
<td>2.63158</td>
<td>75.09149</td>
</tr>
<tr>
<td>GINI_10</td>
<td>GINI index of income disparity, 2010</td>
<td>0.45547</td>
<td>0.0348</td>
<td>0.3321</td>
<td>0.5985</td>
</tr>
<tr>
<td>xpov10</td>
<td>estimated proportion of people of all ages in poverty for income year, 2010</td>
<td>0.14062</td>
<td>0.05214</td>
<td>0.03408</td>
<td>0.53492</td>
</tr>
</tbody>
</table>
## Table 2. Results of Fitting Multiple Regression Models for Predictors of Breast Cancer Rates (n=2430)

<table>
<thead>
<tr>
<th></th>
<th>Model 1 - OLS in GeoDaSpace</th>
<th>Model 2 – Spatial Error Model (HET)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.</td>
<td>S.E.</td>
</tr>
<tr>
<td>Intercept</td>
<td>23.212</td>
<td>6.123</td>
</tr>
<tr>
<td>Divr10</td>
<td>-23.890</td>
<td>5.330</td>
</tr>
<tr>
<td>G25UP1014</td>
<td>0.258</td>
<td>0.052</td>
</tr>
<tr>
<td>GINI_10</td>
<td>67.729</td>
<td>14.413</td>
</tr>
<tr>
<td>Iblack10</td>
<td>32.221</td>
<td>2.718</td>
</tr>
<tr>
<td>Iwhite10</td>
<td>45.941</td>
<td>3.002</td>
</tr>
<tr>
<td>xpov10</td>
<td>-68.027</td>
<td>9.023</td>
</tr>
<tr>
<td>lambda</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.1947</td>
<td>-</td>
</tr>
<tr>
<td>Model F-test</td>
<td>97.61</td>
<td>-</td>
</tr>
<tr>
<td>(df1, df2)</td>
<td>(7, 2423)</td>
<td>-</td>
</tr>
</tbody>
</table>

### Spatial Dependence Diagnostics

<table>
<thead>
<tr>
<th></th>
<th>MI/DF</th>
<th>Est.</th>
<th>$p$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jarque-Bera Test</td>
<td>2</td>
<td>2195.771</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Bruesch-Pagan test</td>
<td>6</td>
<td>55.82</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Moran's I (error)</td>
<td>0.223</td>
<td>18.096</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Lagrange Multiplier (lag)</td>
<td>1</td>
<td>296.865</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Robust LM (lag)</td>
<td>1</td>
<td>2.836</td>
<td>0.0922</td>
</tr>
<tr>
<td>Lagrange Multiplier (error)</td>
<td>1</td>
<td>320.585</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Robust LM (error)</td>
<td>1</td>
<td>26.556</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Est= Parameter Estimate; S.E.= standard error; lambda = Nuisance Parameter; HET = KP Heteroscedastic Errors; Multicollinearity Condition Number = 53.816
Table 3. Results of Fitting Multiple Regression Models for Predictors of Triple Negative Breast Cancer Rates (n=2430)

<table>
<thead>
<tr>
<th></th>
<th>Model 1 - OLS in GeoDaSpace</th>
<th>Model 2 - Spatial Lag Model</th>
<th>Model 3 - Spatial Lag + Error (HET)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.</td>
<td>S.E.</td>
<td>p value</td>
</tr>
<tr>
<td>Intercept</td>
<td>5.066</td>
<td>1.277</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Divr10</td>
<td>-5.882</td>
<td>1.112</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>G25UP1014</td>
<td>-0.009</td>
<td>0.011</td>
<td>0.405</td>
</tr>
<tr>
<td>GINI_10</td>
<td>-4.645</td>
<td>3.007</td>
<td>0.123</td>
</tr>
<tr>
<td>Iblack10</td>
<td>9.975</td>
<td>0.567</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Iwhite10</td>
<td>4.928</td>
<td>0.626</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>xpov10</td>
<td>0.299</td>
<td>1.882</td>
<td>0.874</td>
</tr>
<tr>
<td>rho</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>lambda</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>R2</td>
<td>0.1196</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Model F-test</td>
<td>54.85</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(df1, df2)</td>
<td>(7, 2423)</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Spatial Dependence Diagnostics

<table>
<thead>
<tr>
<th></th>
<th>MI/DF</th>
<th>Est.</th>
<th>p value</th>
<th>MI/DF</th>
<th>Est.</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jarque-Bera Test</td>
<td>2</td>
<td>1251.63</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bruesch-Pagan test</td>
<td>6</td>
<td>163.652</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moran's I (error)</td>
<td>0.1214</td>
<td>9.918</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lagrange Multiplier (lag)</td>
<td>1</td>
<td>108.189</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lagrange Multiplier (error)</td>
<td>1</td>
<td>95.219</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Robust LM (lag)</td>
<td>1</td>
<td>13.881</td>
<td>0.0002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Robust LM (error)</td>
<td>1</td>
<td>0.912</td>
<td>0.3397</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anselin-Kelejian Test</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>9.983</td>
<td>.0016</td>
</tr>
</tbody>
</table>

Est= Parameter Estimate; S.E.= standard error; rho = Autoregressive Parameter; HET = KP Heteroscedastic Errors; Multicollinearity Condition Number = 53.816
References


https://datawarehouse.hrsa.gov/data/dataDownload.aspx


RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER


Mobley, L. R., & Kuo, T.-M. (2015). Geographic and Demographic Disparities in Late-stage Breast and Colorectal Cancer Diagnoses Across the US. *AIMS public health, 2*(3), 583-600. doi:10.3934/publichealth.2015.3.583


Chapter 4
Research Study 3

Manuscript 3

Multi-level Analysis of Person-, County- and State-Level Contributors to Triple-Negative Breast Cancer Diagnosis among Women in the United States

Lia C Scott¹, MPH, Lee Mobley¹, PhD, Tzy-Mey Kuo², PhD, Dora Il’yasova¹, PhD

¹Georgia State University, School of Public Health, Atlanta, GA
²University of North Carolina, Lineberger Cancer Center, Chapel Hill, NC

Correspondence:
Lia C Scott, MPH
Georgia State University
School of Public Health
PO Box 3995
Atlanta, GA 30303
Lscott19@gsu.edu

Manuscript Word Count:
Manuscript Figure Count: 1
Manuscript Table Count: 2
RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER

Abstract

While multiple individual level factors have been identified to play a role in the etiology of breast cancer and triple negative breast cancer, few studies have used mixed modeling techniques to explore the role that additional levels of predictors may play in triple-negative breast cancer. Since disparities persist even when known factors are accounted for, the scope of research must be expanded to examine factors that contribute to disparate outcomes at an ecological level. Mixed modeling can perhaps better account for the spatial heterogeneity in these cases. Random intercept mixed models can account for different effects across areas, yet it assumes higher levels are independent, thus is limited in terms of handling spatial dependence. In this study, we explored the odds of triple negative diagnosis, given breast cancer diagnosis, at the individual level controlling for individual, county and state level variables. When controlling for county and state level predictors, disparities by age, race and stage persisted. Non-Hispanic black women had consistently had twice the odds of diagnosis with TNBC, women age 40 and under had 1.7 times the odds of diagnosis and women diagnosed at late stage had 1.5 times the odds of diagnosis. County-level residential segregation and educational attainment variables were significant predictors of triple negative diagnosis, while no state level policy variables were statistically significant predictors, after controlling statistically for random state intercepts which account for other omitted state variables. Residential isolation proved to be disadvantageous to diagnosis, while residential diversity and area educational attainment were protective. Future studies should continue to explore various environmental factors, physical and social, that contribute the variation in disparate rates of diagnosis.
Introduction

The use of the mixed models is gaining popularity in studies that investigate area-level risk factors (e.g. socioeconomic conditions, residential segregation, social capital, access to primary care healthcare providers, and state insurance mandates) related to individual breast cancer outcomes (Gomez et al., 2015). Mixed models can provide greater insights and explain substantially more of the variation in the observed TNBC outcomes. Individual risk factors are consistently studied and identified, and neighborhood deprivation has been associated with poor breast cancer diagnosis, treatment and survival (Downing, Prakash, Gilthorpe, Mikeljevic, & Forman, 2007; Akinyemiju et al., 2013; Tannenbaum, Koru-Sengul, Miao, & Byrne, 2013; Markossian, Hines, & Bayakly, 2014; Thomson, Hole, Twelves, Brewster, & Black, 2001; Sprague et al., 2011). No studies, to our knowledge, have examined multi-level effects, specifically of community deprivation, on triple negative breast cancer (TNBC) diagnosis. The inclusion of county and state level variables in multilevel modeling can potentially explain additional variation in TNBC diagnosis at the individual level. With this enhanced knowledge, and effective dissemination and translation of it, policies and interventions can be designed and targeted to address the barriers and gaps that contribute to the observed TNBC outcomes. Model building for multilevel analysis will allow us to explore how these various levels of predictors contribute to diagnosis of TNBC among BC patients.

The literature using atomistic modeling of single-level individual data has provided several facts. One study found that having low self-reported education was associated with subtypes of estrogen receptor negative and progesterone receptor negative breast cancers (Trivers et al., 2009). Additionally, women in their study with triple negative tumors were more often of lower self-reported SES. Low socioeconomic status has been linked to decreased rates
RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER

of screening, greater probability for late-stage diagnosis, receipt on inadequate and disparate treatment and higher mortality from breast cancer (Bigby & Holmes, 2005; Gerend & Pai, 2008). Regardless of race, poverty is associated with poorer breast cancer outcomes (Gerend & Pai, 2008). Women who reside in disadvantaged communities may be required to travel longer distances with longer wait times to utilize screening and treatment facilities. These factors can cause a major hindrance in regular physician visits (Mandelblatt, Andrews, Kao, Wallace, & Kerner, 2010; Mandelblatt, Andrews, Kerner, Zauber, & Burnett, 1991; Mandelblatt, Yabroff, & Kerner, 1999; Vernon, Vogel, Halabi, & Bondy, 1993).

Key factors in health care utilization involve accessibility. There is evidence to suggest that when nurse practitioners are allowed to practice and prescribe medicine independently, that adults utilize the healthcare system more frequently (Kuo, Loresto, Rounds, & Goodwin, 2013; Mobley et al., 2017; Stange, 2014). Underinsurance takes into account an insured adult’s reported out-of-pocket costs over the course of a year, not including premiums and deductibles. A person is defined as someone whose (a) out-of-pocket costs, excluding premiums, over the prior 12 months are equal to 10 percent or more of household income; (b) out-of-pocket costs, excluding premiums, are equal to 5 percent or more of household income if income is under 200 percent of the federal poverty level ($22,980 for an individual and $47,100 for a family of four); or (c) deductible is 5 percent or more of household income (Collins, Rasmussen, Beutel, & Doty, 2015). From 2010 to 2014, twenty-three percent of insured adults were underinsured. Uninsured and underinsured women are less likely to undergo screening, more likely to receive a late stage diagnosis and less likely to survive (Bradley, Given, & Roberts, 2002; Buseman, Byers, Finch, & Jacobellis, 2002; Gerend & Pai, 2008; Gordon, Rundall, & Parker, 1998; Hsia et al., 2000; Roetzheim et al., 1999). Low-income women have mammography screening rates that are lower
RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER

than higher income women (Peek & Han, 2004). Additionally, women with lower education have lower mammography screening rates (Kerner et al., 2001).

Few studies have explored the role of social environmental risk factors in triple negative breast cancer diagnosis and, of those, most are analyzed at a single level. Chu and colleagues found that race or ethnicity had no effect on recurrence or survival among TNBC patients, when controlling for ZIP code-based income, but did not use hierarchical techniques (Chu, Henderson, Ampil, & Li, 2012). Parise and Caggiano (2017) created an SES index with Census data and did not find an increase in odds of TNBC compared to ER-/PR-/HER2+ due to low SES for all races/ethnicities using logistic regression. Bauer and colleagues use a similar SES index as above in logistic regression and found that TNBC affects younger, non-Hispanic black and Hispanic women in areas of low SES (Bauer et al, 2007).

This paper explores the relationship between age, race and stage at diagnosis and triple-negative diagnosis among women with breast cancer. As noted in the literature above, disparities based on age, race, and stage have been widely published, but additional contributions to the literature can be made by exploring and controlling for area-level contextual effects (Howlader et al., 2013). In this paper, we include additional social environmental risk factors that may better inform both policy and clinical practice. We include area-level factors including residential segregation, income inequality, educational attainment and social capital (at the county-level) and health insurance rates, mammography screening, and scope of practice regulation policy (at the state level). The social environment variables in the socioecological model on the cancer continuum drives the theory behind inclusion of variables in this study (Figure 1) (Gomez et al, 2015). We may still find that there are certain disparities that persist with the addition of these variables. The use of multilevel modeling allows you to control for these contextual factors and
RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER

examine whether the aforementioned person-level relationships remain robust after inclusion of additional levels of context. We are also able to examine whether there is any substantial impact from area context by looking at the level-2 and level-3 results.

The research questions are as follows: What is the predicted probability of triple negative breast cancer diagnosis, given breast cancer diagnosis, among females in an average county in the United States? What community variables are strong predictors of triple negative breast cancer diagnosis among breast cancer cases? When we control for patient, county and state characteristics, what is the relationship between patient race and odds of TNBC diagnosis, given breast cancer diagnosis? The research hypotheses are: TNBC diagnosis will vary for females across counties and counties within states and will be associated with several factors reflecting community disadvantage. We anticipate that the odds of triple negative breast cancer diagnosis will remain higher for Non-Hispanic black women compared to Non-Hispanic white women, even when controlling for county and state level characteristics. Multilevel analysis will explain the variation in diagnosis within and between racial groups using person, county and state level predictors.

**Methods**

We examined all breast cancer cases diagnosed during 2010–2014 from the United States Cancer Statistics (USCS) database, which is a population-based surveillance system of cancer registries with data representing 99% of the U.S. population (Richardson et al, 2016). Most states participate in the USCS registry data system, but five did not provide county-level breast cancer data – Illinois, Kansas, Michigan, Minnesota and Missouri – and four did not code for triple negative data – Connecticut, Iowa, New Mexico, and Utah. Alaska and Hawaii were excluded from analysis due to missing contextual data and ill-defined counties, leaving 39 states in the
analysis. Community characteristics will be represented using data from the Spatial Impact Factor Database (Mobley, 2015). The database was supplemented by additional data from the Area Health Resource File and Pennsylvania State University (Bureau of Health Workforce, 2017; Rupasingha, Goetz, & Freshwater, 2006).

Person-Level Factors

Late stage was defined as diagnosis at stage 3 and beyond. Age groups were defined as less than 40, 40 – 49, 50 -64, 65 – 74, and 75 and older with age 50 -64 serving as the referent group. There were six race categories in the study: Non-Hispanic white, Hispanic, Non-Hispanic black, American Indian/Alaskan Native, Asian, and Other, with Non-Hispanic white serving as the referent category.

County-level factors

County level data included in the model were non-Hispanic black and non-Hispanic white isolation indices, Theil’s diversity index, Gini index of income disparity, poverty, social capital and educational attainment. Non-Hispanic black and non-Hispanic white isolation indices are measures of exposure that indicate the probability the specified race group will encounter another person from their race group within the areal unit (county) (Massey & Denton, 1988). A value of 1 indicates a perfectly racially isolated county for the respective race. Theil’s Diversity index is a multi-group measure that reflects the level of diversity within the county. Counties where tracts all have the same composition (all individuals in a population are associated with the same racial group) have a low diversity index while counties where different races or ethnicities are separated into cultural enclaves among the tracts will have high indices (Mobley, 2015; White, 1986). A value of 0 for the diversity index indicates no diversity in the population while a value of 1 indicates maximum diversity, where individuals are evenly distributed among
two or more mutually exclusive groups (Roberto, 2015). The Gini coefficient is the most commonly used indicator of income inequality, and its use here allows for comparisons with other income inequality studies (Farley, 2006; Jones-Smith, Gordon-Larsen, Siddiqi, & Popkin, 2011). The GINI Index of family income inequality reflects the degree to which income is equally distributed between families within a county. A coefficient of 0 represents a perfectly equal society where every family has the same income, while a coefficient of 1 represents a perfectly unequal society with a great divergence between the richest and poorest families. Poverty was represented as the proportion of the population in poverty as defined by the U.S. Census. Social capital was a normalized index developed using the number of the following establishments in each county: (a) civic organizations; (b) bowling centers; (c) golf clubs; (d) fitness centers; (e) sports organizations; (f) religious organizations; (g) political organizations; (h) labor organizations; (i) business organizations; and (j) professional organizations, along with the percentage of voters in presidential elections, county-level Census response rate, and the number of tax-exempt non-profit organizations (Alesina & La Ferrara, 2000; Knack, 2002; Rupasingha et al., 2006). Education was further defined from US Census data as the percent of the population age 25 and older with a graduate or professional degree (Evenden, Harper, Brailsford, & Harindra, 2006).

State-level factors

State level variables were nurse practitioner (NP) autonomy laws, percent underinsured, and mammography screening rates. Data on NP regulations was provided by the National Conference of State Legislatures (NCSL). Underinsurance data are presented as percent of adults who were underinsured in 2012 (Collins et al., 2015). Mammography screening rates were obtained from the Centers for Disease Control and Prevention Behavioral Risk Factor
RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER

Surveillance System. These values represent the percent of women age 40 and over who had a mammogram in 2010.

The dataset was combined and summary statistics were computed using SAS Software (SAS 9.4, SAS Institute Inc., Cary, NC). Triple-negative diagnosis was coded using site specific factors 1, 2 and 15. County and state level data were merged using the FIPS codes. Univariate and bivariate relationships were examined in SAS 9.4. Due to the hierarchical structure of the data, with cases nested in counties, nested in states, we employed mixed modeling techniques. The outcome of interest was triple negative diagnosis (Yes/No) at the individual. In all the mixed models, we specified a binary distribution with a logit link. Utilization of the logit response function is appropriate when the response probability is small because the logistic distribution has greater tail probability than the normal distribution (Gibbons & Hedeker, 1997). Due to the size of the dataset and limitations in time, we optimized PROC GLIMMIX to ensure convergence in the models. We suppressed the print options and random effects solutions. The denominator degrees of freedom method employed was between-within, which utilizes less computing resources than the default containment method. Newton-Raphson Ridging option was employed with max iterations set to 50, as this option is appropriate in binary distributions (Kiernan, Tao, & Gibbs, 2012). Laplacian approximation was not employed in this study due to non-convergence of models, rather residual subject-specific pseudo-likelihood was used. This does not allow for model comparison using -2 Log Likelihood values and model output did not include any other model fit statistics such as the AIC or BIC.

An intercept-only mixed model, with a test for random intercepts at level 2 and 3, was run and the intraclass correlation was calculated (Snijders & Bosker, 2011). The ICC explains the percent of the total variance accounted for at various levels. The ICC for state level was
RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER

\[ \frac{\sigma^2(3)}{(\sigma^2(2)+\sigma^2(3)+\pi^2/3)} \] , the ICC at county within state level was \[ \frac{\sigma^2(3)}{(\sigma^2(2)+\sigma^2(3))} \] , and the ICC for county and state level was \[ \frac{\sigma^2(2)+\sigma^2(3)}{(\sigma^2(2)+\sigma^2(3)+\pi^2/3)} \] (Raman and Hedeker, 2005). To explore and control for contextual county-level effects, we continued to add on to the intercept-only model. Further model building included a model with level-1 predictors followed by a model with level-2 predictors, followed by three separate models, one for each level-3 predictor. We allowed for random effects at the county and state level. A small ICC indicated that the variance in the observed response stems more from individual differences within groups. However, omitted contextual effects could conceivably impart bias on the included person-level effects. We converted the estimates to population case-only odds ratios. We refer to this odds ratio with this nomenclature because the population contains only cases on breast cancer, thus it is not a true case-control. This population case-only odds ratio approach has been used in cancer studies where the authors are examining genes and mutations that lead to cancer (Rosenbaum, 2004).

Results

This study included 1,102,113 individuals, 2430 counties, and 39 states. Descriptive statistics for the independent variables are presented in Table 1. The intraclass correlation was 0.008 for states, 0.019 for counties and states, and 0.427 for counties. This indicates that the likeness of cases in the same states is 0.81%, the likeness of cases in the same counties in the same states is 1.89%, and the likeness of counties in the same states is 42.7%. The individual predicted probability of diagnosis in an average county within an average state is 8.7%, derived from the log odds of TNBC diagnosis in the intercept only model (Table 2).

The results of model-building are presented in Table 2. Age, race and stage were significant predictors of TNBC diagnosis, given BC diagnosis, in all 6 models. In terms of race,
RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER

the referent group was non-Hispanic white women in the case only population. Non-Hispanic black women had twice the odds of diagnosis even when controlling for county and state level covariates. Hispanic and American Indian/Alaskan Native women also had higher odds of diagnosis. Asian and Other women had lower odds of diagnosis. Those age less than 40 had the highest odds of diagnosis compared to the referent group of 50-64 years. This age group was 1.7 times more likely to be diagnosed with TNBC given BC diagnosis. Age 40-49 had slightly higher odds, while both age groups above 65 had lower odds of diagnosis. Those diagnosed at late stage had 1.6 times the odds of diagnosis with TNBC compared to those diagnosed prior to stage III.

Isolation indices, the diversity index, and education were statistically significant predictors of TNBC diagnosis given BC diagnosis. Compared to a perfectly integrated county, an individual in a perfectly segregated county has approximately 1.3 times the predicted odds of diagnosis with TNBC in regard to non-Hispanic black isolation, and approximately 1.2 times the predicted odds of diagnosis in regard to non-Hispanic white isolation. Compared to a perfectly isolated county (where all tracts have the same composition), individuals in a county with maximum diversity (multi-group) have approximately 0.7 times the predicted odds of diagnosis with TNBC. For every one unit increase in percent of the population that is educated beyond college, the predicted odds of diagnosis with TNBC is decreased by 0.9. No state level variables were statistically significant predictors of triple negative diagnosis.

Discussion

Even with small variance components, it is still imperative to explore social and policy factors that may have implications in individual TNBC diagnosis. In an average county and state, the individual level predicted probability of TNBC diagnosis was 8.7%, given breast cancer
RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER

diagnosis. Even when controlling for county and state level factors, non-Hispanic black women experienced higher odds of diagnosis of TNBC, followed by Hispanic and American Indian/Alaskan Native women. Asian women and Other women had lower odds of diagnosis. Younger age and late stage at diagnosis also conferred higher odds of diagnosis.

At the county level, residential segregation is associated with TNBC diagnosis. Non-Hispanic black and non-Hispanic white isolation confers higher odds of diagnosis, regardless of individual race, controlling for all other factors. However, the diversity index confers lower odds of diagnosis. The diversity index is a multi-group measure of spatial evenness that quantifies how evenly racial groups are distributed across the tracts within each county. A value of 0 for the diversity index indicates no diversity, where all tracts have the same composition, while a value of 1 indicates maximum diversity, where all racial or ethnic groups are evenly distributed but may be segregated into cultural enclaves among tracts (Roberto, 2015). Thus, diversity within the county confers lower odds of diagnosis at the individual level. Residential segregation has been a widely studied phenomenon with no consensus on whether it is detrimental or beneficial to health. The aim is to determine whether it creates an adverse environment or provides social support (Acevedo-Garcia, Lochner, Osypuk, & Subramanian, 2003; Boustan, 2013; Bower et al., 2015; Charles, 2003; Collins, 1999; Collins & Williams, 1999; Dai, 2010; Dinwiddie, Gaskin, Chan, Norrington, & McCleary, 2013; Haas et al., 2008a, 2008b; Hao et al., 2011; Hayanga, Zeliadt, & Backhus, 2013; Kershaw, Albrecht, & Carnethon, 2013; Mobley, Kuo, Scott, Rutherford, & Bose, 2017; Mobley, Scott, Rutherford, & Kuo, 2016; Williams & Collins, 2001).

For these particular outcomes, we found that area-level residential segregation, specifically, isolation, was detrimental for health, and that diversity within counties provided better outcomes.
Area-level higher education conferred lower odds of individual diagnosis while income inequality and poverty had non-significant effects.

No state level variables were significant predictors of individual level TNBC diagnosis. However, we expected nurse practitioner laws to have an impact on diagnosis as the literature demonstrates improvements in the provision of preventive services to patients when a nurse or NP offered screening directly to the patients, bypassing the physician (Beach et al, 2006). There is also favorable evidence to support the use of bypassing providers of minority patients to offer standardized services (Beach et al, 2006). Another study on Medicare beneficiaries found that NPs are also more likely to serve younger populations, females, and minorities, as well as dually eligible Medicare and Medicaid. Importantly, they were more likely to practice in primary care shortage areas (DesRoches, Gaudet, Perloff, Donelan, Iezzoni & Buerhaus, 2013). We may not have found a significant effect because the variable is analyzed at three levels. Nurse practitioner autonomy could potentially increase access for communities most likely to be affected by TNBC diagnosis, specifically younger female populations. Future studies where the variable is dichotomized as full autonomy versus reduced or restricted practice may elucidate the role that state-level nurse practitioner autonomy plays in individual diagnosis. From 1987 – 2000, 42 stated adopted screening mandates requiring private insurers within the state to include mammography benefits in insurance plans (Bitler and Carpenter, 2011). Given the screening mandate, we would expect states to have a higher utilization of mammography screening which in turn would affect outcomes. However, there isn’t much literature to show that generous insurance coverage will increase screening utilization and millions of Americans still remain un- or underinsured (Bitler and Carpenter, 2011). Overall, we expected access to care, represented by the state variables, to lower the odds of diagnosis with TNBC, rather than finding a null
RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER

effect. However, it is important to note that even with control of these state-level policy variables, disparities still existed and individual and county level effects were not attenuated.

It is important to note that penalized quasi-likelihood estimates of random components in generalized mixed models with dichotomous outcomes have previously been found to have a downward bias when there are few observations per group (Capanu, Gönen, & Begg, 2013; Goldstein & Rasbash, 1996; Raudenbush, Yang, & Yosef, 2000; Rodriguez & Goldman, 1995). Given the size of the dataset, we were not concerned with small cell sizes, but acknowledge this as a potential source of bias. Future studies should consider the use of a numerical integration method such as Laplacian or Adaptive Quadrature. It is important to consider the role the social environment plays in individual level diagnosis. Further exploration of macro-level factors is necessary to truly evaluate disparities in triple negative diagnosis. Individual level disparities were not eliminated with the addition of county and state level variables. We found that even when we control for individual age and stage, county residential segregation, education, income inequality levels, and state level policy variables, Non-Hispanic black women still had the highest odds of TNBC diagnosis compared to non-Hispanic white women, followed by Hispanic and AI/AN women. The addition of SES variables at the individual level as well as cross-level interactions may provide different insights into what is driving disparities. An exploration of the physical context (e.g. environmental contaminants) may elucidate further what may be driving the observed TNBC disparities, as well as policy variables.
Figure 1. Socioecological model on the Cancer Continuum (Gomez et al., 2015)

Table 1. Descriptive Statistics for County and State Level Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Variable Description</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iblack10</td>
<td>Isolation Index: probability that non-Hispanic black persons will meet other non-Hispanic black persons within defined area, 2010</td>
<td>0.26251</td>
<td>0.20741</td>
<td>0</td>
<td>0.8647</td>
</tr>
<tr>
<td>Iwhite10</td>
<td>Isolation Index: probability that non-Hispanic white persons will meet other non-Hispanic white persons within defined area, 2010</td>
<td>0.73452</td>
<td>0.15163</td>
<td>0.03154</td>
<td>1</td>
</tr>
<tr>
<td>Divr10</td>
<td>Diversity Index (Theil Index): measures the even-ness or uneven-ness of the spatial distribution of population subgroups in tracts within areas (counties), 2010</td>
<td>0.1981</td>
<td>0.0916</td>
<td>0</td>
<td>0.69685</td>
</tr>
<tr>
<td>G25UP1014</td>
<td>percent of population aged 25+ years with a graduate or professional degree, 2010-2014</td>
<td>29.56426</td>
<td>10.65997</td>
<td>2.63158</td>
<td>75.09149</td>
</tr>
<tr>
<td>GINI_10</td>
<td>GINI index of income disparity, 2010</td>
<td>0.45547</td>
<td>0.0348</td>
<td>0.3321</td>
<td>0.5985</td>
</tr>
<tr>
<td>sk09</td>
<td>Social capital index created by Pennsylvania State University, 2009</td>
<td>-0.70851</td>
<td>0.83828</td>
<td>-3.92523</td>
<td>17.4405</td>
</tr>
<tr>
<td>x pov10</td>
<td>estimated proportion of people of all ages in poverty for income year 2010</td>
<td>0.14062</td>
<td>0.05214</td>
<td>0.03408</td>
<td>0.53492</td>
</tr>
<tr>
<td>UNDER12</td>
<td>percent of the population that is underserved by primary care providers, 2012 (Kaiser)</td>
<td>10.37237</td>
<td>5.20211</td>
<td>0.32525</td>
<td>31.50783</td>
</tr>
<tr>
<td>XMAM10</td>
<td>percent of the population that received mammograms in 2010 (BRFSS)</td>
<td>78.57239</td>
<td>3.71165</td>
<td>68.3</td>
<td>87.5</td>
</tr>
<tr>
<td>NP2012</td>
<td>Whether state allowed NPs autonomy to diagnose, treat, and prescribe without physician oversight. 0: Full autonomy; 1: Reduced Practice; 2: Restricted Practice</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NP2012 = 0 (County)</td>
<td>1575</td>
<td>455</td>
<td>400</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NP2012 = 1 (County)</td>
<td>18</td>
<td>7</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NP2012 = 2 (State)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Racial and Geospatial Disparities in Triple-Negative Breast Cancer

Table 2. Mixed Modeling Results for Three-Level Binary Outcome Models to Predict Triple-Negative Diagnosis among Females with Breast Cancer

<table>
<thead>
<tr>
<th>Model</th>
<th>Intercept Only</th>
<th>Level 1</th>
<th>Level 1 + 2</th>
<th>Level 1 + 2 + NP Law</th>
<th>Level 1 + 2 + Under</th>
<th>Level 1 + 2 + Mamm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>Estimate</td>
<td>SE</td>
<td>OR</td>
<td>95% CI</td>
<td>Estimate</td>
<td>SE</td>
</tr>
<tr>
<td>Model 1</td>
<td>Intercept Only</td>
<td>-2.3546</td>
<td>0.02789</td>
<td>-2.5285</td>
<td>0.02036</td>
<td>-2.4953</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic white</td>
<td>REF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic black</td>
<td>2.068</td>
<td>2.029, 2.107</td>
<td>2.06</td>
<td>2.02, 2.101</td>
<td>2.06</td>
<td>2.02, 2.101</td>
</tr>
<tr>
<td>Hispanic</td>
<td>1.186</td>
<td>1.155, 1.218</td>
<td>1.189</td>
<td>1.158, 1.222</td>
<td>1.189</td>
<td>1.158, 1.222</td>
</tr>
<tr>
<td>AI/AN</td>
<td>1.116</td>
<td>1.014, 1.229</td>
<td>1.114</td>
<td>1.011, 1.227</td>
<td>1.113</td>
<td>1.011, 1.227</td>
</tr>
<tr>
<td>Asian</td>
<td>0.897</td>
<td>0.862, 0.934</td>
<td>0.903</td>
<td>0.868, 0.94</td>
<td>0.903</td>
<td>0.868, 0.94</td>
</tr>
<tr>
<td>Other</td>
<td>0.754</td>
<td>0.682, 0.833</td>
<td>0.756</td>
<td>0.684, 0.835</td>
<td>0.756</td>
<td>0.684, 0.835</td>
</tr>
<tr>
<td>Age Groups</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50-64</td>
<td>REF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;40</td>
<td>1.769</td>
<td>1.72, 1.819</td>
<td>1.773</td>
<td>1.724, 1.823</td>
<td>1.773</td>
<td>1.724, 1.823</td>
</tr>
<tr>
<td>40-49</td>
<td>1.076</td>
<td>1.055, 1.097</td>
<td>1.078</td>
<td>1.057, 1.099</td>
<td>1.078</td>
<td>1.057, 1.099</td>
</tr>
<tr>
<td>65-74</td>
<td>0.845</td>
<td>0.83, 0.861</td>
<td>0.844</td>
<td>0.829, 0.86</td>
<td>0.844</td>
<td>0.829, 0.86</td>
</tr>
<tr>
<td>75+</td>
<td>0.819</td>
<td>0.802, 0.836</td>
<td>0.818</td>
<td>0.801, 0.835</td>
<td>0.818</td>
<td>0.801, 0.835</td>
</tr>
<tr>
<td>Late Stage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>REF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1.565</td>
<td>1.543, 1.588</td>
<td>1.563</td>
<td>1.54, 1.586</td>
<td>1.563</td>
<td>1.54, 1.586</td>
</tr>
<tr>
<td>NP2012</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>REF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.065</td>
<td>0.97, 1.17</td>
<td>1.065</td>
<td>0.97, 1.17</td>
<td>1.065</td>
<td>0.97, 1.17</td>
</tr>
<tr>
<td>2</td>
<td>1.026</td>
<td>0.946, 1.112</td>
<td>1.026</td>
<td>0.946, 1.112</td>
<td>1.026</td>
<td>0.946, 1.112</td>
</tr>
<tr>
<td>UNDER12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XMAM10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Variance Components

<table>
<thead>
<tr>
<th>Variance Components</th>
<th>Estimate</th>
<th>SE</th>
<th>Estimate</th>
<th>SE</th>
<th>Estimate</th>
<th>SE</th>
<th>Estimate</th>
<th>SE</th>
<th>Estimate</th>
<th>SE</th>
<th>Estimate</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variance(State)</td>
<td>0.0272</td>
<td>0.0068</td>
<td>0.0126</td>
<td>0.0033</td>
<td>0.0092</td>
<td>0.0024</td>
<td>0.0092</td>
<td>0.0025</td>
<td>0.0095</td>
<td>0.0025</td>
<td>0.0095</td>
<td>0.0025</td>
</tr>
<tr>
<td>Variance(County)</td>
<td>0.0365</td>
<td>0.0026</td>
<td>0.0229</td>
<td>0.0021</td>
<td>0.0173</td>
<td>0.0018</td>
<td>0.0173</td>
<td>0.0018</td>
<td>0.0173</td>
<td>0.0018</td>
<td>0.0173</td>
<td>0.0018</td>
</tr>
</tbody>
</table>
RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER

References


Bradley, C. J., Given, C. W., & Roberts, C. (2002). Race, Socioeconomic Status, and Breast Cancer Treatment and Survival. *Journal of the National Cancer Institute, 94*(7), 490-496. doi:10.1093/jnci/94.7.490


RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER


RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER


RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER


Kuo, Y. F., Loresto, F. L., Jr., Rounds, L. R., & Goodwin, J. S. (2013). States with the least restrictive regulations experienced the largest increase in patients seen by nurse practitioners. *Health Aff (Millwood), 32*(7), 1236-1243. doi:10.1377/hlthaff.2013.0072


Mobley, L. R., & Kuo, T.-M. (2015). Geographic and Demographic Disparities in Late-stage Breast and Colorectal Cancer Diagnoses Across the US. *AIMS public health, 2*(3), 583-600. doi:10.3934/publichealth.2015.3.583


RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER


Chapter 5

Dissertation Summary and Future Directions in Research
Previous studies in the literature have focused on individual level risk factors, especially with the triple negative subtype. The literature found that triple negative cases account for 10% to 15% of all breast cancer cases, while this study found that approximately 8% of cases were triple negative using site specific factors 1, 2 and 15. Several (13) states were excluded from analysis due to lack of either data as a whole or triple negative classification data. Also, the study began with 2010 data because prior to that time, registries inconsistently coded TNBC. There is a need for more comprehensive and reliable data collection from registries, in order to truly capture each case. These studies in this dissertation found that race, age and stage at diagnosis play a significant role, regardless of additional control variables.

The three manuscripts presented in this dissertation focused on identifying individual and community level factors in triple-negative diagnosis and geographic distribution and disparities in breast cancer and triple negative breast cancer rates. The overarching goal of this study was to evaluate racial and geographic disparities in triple-negative breast cancer (TNBC) diagnoses by race or ethnicity both within regions and across regions and advance the field of population-based research in breast cancer disparities through innovative statistical techniques. Goals of each of the three individual manuscripts were, respectively: (1) to establish the odds of diagnosis of triple negative breast cancer among breast cancer cases by age, race, and stage at diagnosis, (2) evaluate the geospatial distribution of county-level breast and triple negative breast cancer rates and their associations with community level characteristics (socioeconomic conditions and, residential segregation), and (3) examine multilevel effects of county and state level variables on individual level triple negative diagnosis among breast cancer cases.

The literature demonstrates that TNBC accounts for 12-15% of all breast cancers, but no studies have been conducted at a national level with as rich of data as the USCS. This study
RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER

identified 96,749 TNBC cases among 1,151,724 BC cases in the United States across 5 years, 2010 – 2014. The first study aimed to validate previous findings in the literature. No studies have been conducted on TNBC at the national level with as comprehensive of a dataset, and this is the first of its kind to explore differences in person-level factors at such a large scale. We utilized traditional epidemiologic methods to evaluate individual level differences, but these methods can confound race and place. The second study built upon the first to describe geospatial patterns in the disease at the county-level using exploratory spatial data analysis and examine associations with contextual factors using confirmatory spatial data analysis. There are no studies in the literature that examine TNBC and BC outcomes and clustering patterns across the United States, nor use spatial regression to account for spatial dependence in the outcomes. We know that there are both micro- and macro-social processes that can impact health and health outcomes. In the third study, we aim to account for multi-level influences on individual TNBC diagnosis, to address the fallacies of single-level research.

The first study found that triple negative breast cancer cases account for fewer breast cancer cases than previously described in the literature. This study found that TNBC accounted for 8.5% of cases, whereas previous literature estimates that TNBC accounts for 12-15%. The study found that, consistent with the literature, in both unadjusted and adjusted analyses, race, age and stage were significant predictors of triple negative diagnosis. However, this study found that the burdens among women of color, specifically non-Hispanic black and Hispanic women, younger women, and women diagnosed at later stage are actually higher than previously estimated. We found evidence of different distribution of age, race and stage at diagnosis compared to previous studies. Non-Hispanic black and Hispanic women had higher odds of triple-negative diagnosis, compared to Non-Hispanic white women. Women under the age of 50
had higher odds of triple negative diagnosis while women over 64 had lower odds of diagnosis, compared to women age 50 to 64. Diagnosis at stage III and beyond and stage IV and beyond also conferred higher odds of diagnosis.

The second study found evidence of significant spatial autocorrelation in both breast cancer and triple negative breast cancer rates across counties. There were distinct geospatial patterns of the crude BC and TNBC rates. The concurrence of breast cancer and triple negative breast cancer clusters demonstrated areas of need where both breast cancer and triple negative breast cancer incidence rates are higher than average, and where breast cancer rates are lower than average while triple negative rates are higher than average. Special attention needs to be paid to counties within these clusters. Spatial regression revealed strong effects for residential segregation and education, implying that community context may vary across communities along with their triple negative diagnosis rates. Additional studies examining other social and physical environmental characteristics need to be employed to further elucidate the association between environmental factors and triple negative diagnosis.

The third study examines spatial heterogeneity in the likelihood of TNBC diagnosis through the use of mixed modeling. The significant variance components at the county and state level suggest that there is correlation among observations at each of these levels. County-level residential segregation and education variables were found to be significant predictors of triple negative diagnosis at the individual level. No state level variables were statistically significant, but inclusion of these variables allow for control for state-level policy. No individual level effects were attenuated due to the addition of county and state level effects, indicating that robust estimates of disparities exist even with the control of various contextual variables.
Overall, the studies show that disparities exist by age, race and stage even when controlling for spatial heterogeneity and dependence. Controlling for these area-level effects is important when population data span the country and geographic disparities are a possible confounder. Several studies have examined whether it is the race or the place that is determining disparities, and the importance of including these area context variables in population studies that span the country (Gomez et al., 2015). These three individual level factors - age, race and stage - are important to consider in the diagnosis of triple negative breast cancer and targeting of early intervention.

There were several contextual variables that were associated with breast cancer and triple negative breast cancer rates. To discover robust disparity estimates at the individual level, it is necessary to include them in the model for conceptual reasons. Further research should be conducted to identify other contextual variables of interest that may contribute to variation in rates. The addition of compositional effects along with the variables currently included as contextual effects may also reveal different results. As argued by Oakes, to truly know area contextual effects, one must hold constant statistically compositional effects that could confound the estimates due to selection bias of individuals into communities (Oakes, 2009). Because the focus here was on how omitted context might impact individual-level estimates, we used the model specification described here.

Given the large sample size and spatial coverage of the data, these results are more generalizable than previous studies. Triple-negative data did not become reliable within the registries until 2010, and all states still do not include the necessary variables to identify triple negative cases. The study is dependent upon and limited by the accuracy of the registry data. There is a need for exploration of many other variables that may play a role such as food
environment data, since obesity and physical activity have been linked to breast cancer, and physical environment data. Research needs to continue to explore what other variables that may be driving disparities, including the physical environment.

Although the study is the largest of its kind, we were unable to include the full scope of the dataset in the analysis due to data sharing restrictions or lack of information on the TNBC outcome. Triple negative breast cancer can be identified using three site specific factors, SSF1, SSF2, SSF15 equal to ‘020’, and several states did not code appropriately. The study included 39 states of the 48 available. However, we were still able to identify over 1 million BC cases in the 5-year span. We found that triple negative cases accounted for far less than previously identified in the literature. This may be due to the use of a subset of 39 states. Comparisons of findings with the entire SEER registry, as well as annual evaluations of the data, may further validate what number may truly represent the prevalence of TNBC cases among BC cases. In the second study, we utilized crude rates due to lack of appropriate population data within the Research Data Center to create age-adjusted rates. The geospatial portion of this study needs to be repeated and validated using age-adjusted rates. The third study did not find any effects on individual level diagnosis for the state level predictors. We expected nurse practitioner autonomy to have an impact on diagnosis as the literature demonstrated that NPs often serve low access communities. We may not have found an effect because the variable was analyzed at three levels. The mixed modeling could be repeated with the NP variable at 2 levels, either comparing full autonomy to the remaining two categories or full restriction to the remaining two categories. The social capital variable was a normalized variable that was not found to be a significant predictor in any of the mixed models. This creates difficulties in interpretation. In a comparison study of social capital indices, it was found to be a conceptually valid and robust measure of collective social
capital (Lee & Kim, 2013). Future studies could consider manipulation of the variable to a more intuitive measure prior to including it in modeling. Additionally, this study used penalized quasi-likelihood estimates which can have a downward bias in generalized mixed models with dichotomous outcomes and small group cells. Given the size of the dataset, this was not a concern, but it is acknowledged as a potential source of bias. Future studies should employ Laplacian or adaptive quadrature methods. Mixed models account for spatial heterogeneity but not spatial dependence. Future studies should employ spatial mixed modeling techniques as the capability becomes available.

Repeated trend studies from 2010 forward are necessary to track disparities in triple negative breast cancer and determine whether improvements towards equity have been made. Since the study was employed at multiple levels, it can serve to inform policy at both the local and state level. Given the results of this study, there are additional drivers of disparities that are currently unaccounted for. Exploration from a theory driven standpoint can further elucidate these drivers. It is necessary that researchers consider all possible influences on breast cancer disparities and shift focus from the individual to the community. The geospatial analysis showed that there exists geospatial disparities. Future studies could examine geospatial disparities for race-specific or age-specific rates to determine if these patterns persist when the data are stratified.

Overall, this study adds to the growing literature of triple negative breast cancer disparities. It found that triple negative breast cancer cases accounted for fewer breast cancer cases than previously identified in the literature, and that disparities between women based on race, age, and stage at diagnosis were more pronounced. It found that Non-Hispanic black women consistently had twice the odds of diagnosis with triple negative given breast cancer
diagnosis. Women age 40 and under had almost twice the odds of diagnosis and women
diagnosed at stage III and beyond had 1.5 times the odds of diagnosis. The distributions of age,
race and stage at diagnosis were distinctly different from what had been previously identified in
the literature. Additional attention should be paid to other individual level characteristics that
confer a higher risk of diagnosis. It found that residential segregation plays a role in diagnosis as
well. In both the spatial regression and mixed model, the latent variables of isolation indices for
non-Hispanic white and non-Hispanic black persons as well as the diversity index were
consistently statistically significantly associated with TNBC outcomes at the county and
individual level. Residential isolation in non-Hispanic black or non-Hispanic white isolated
counties were associated with an increase in breast cancer or triple negative breast cancer rates,
and increased odds of individual diagnosis with triple negative breast cancer. Diversity within
the counties was associated with a decrease in breast cancer and triple negative breast cancer
rates, and decreased odds of individual diagnosis. County-level higher education was negatively
associated with both breast and triple negative breast cancer rates and conferred lower odds of
triple negative diagnosis. These findings imply there may be additional processes driven by
residential segregation that promulgate the disparity in diagnosis. It is imperative that cancer
research continue to explore the micro- and macro-social processes that impact diagnosis with
TNBC, given its aggressive nature. Treatment development cannot be the only course of action
when tackling such a heterogeneous disease. This study provides a foundation from which
additional studies can move forward from, further exploring individual and contextual factors
that continue to drive disparities in triple negative breast cancer.
RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER

References


RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER


RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER


RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER


analyse risk of hypertension. *Journal of Epidemiology and Community Health, 69*(12), 1199-1207. doi:10.1136/jech-2015-205728


RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER


RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER


Krieger, N., Singh, N., & Waterman, P. D. (2016). Metrics for monitoring cancer inequities: residential segregation, the Index of Concentration at the Extremes (ICE), and breast


doi:10.1016/j.healthplace.2010.11.007

Kuo, Y. F., Loresto, F. L., Jr., Rounds, L. R., & Goodwin, J. S. (2013). States with the least restrictive regulations experienced the largest increase in patients seen by nurse practitioners. Health Aff (Millwood), 32(7), 1236-1243. doi:10.1377/hlthaff.2013.0072


Mobley, L. R., & Kuo, T.-M. (2015). Geographic and Demographic Disparities in Late-stage Breast and Colorectal Cancer Diagnoses Across the US. *AIMS public health, 2*(3), 583-600. doi:10.3934/publichealth.2015.3.583


RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER


RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER


Racial and Geospatial Disparities in Triple Negative Breast Cancer


RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER


RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER


Surveillance, Epidemiology, and End Results (SEER) Program (2018). Number of Person by Race and Hispanic Ethnicity for SEER Participants (2010 Census Data).


RACIAL AND GEOSPATIAL DISPARITIES IN TRIPLE NEGATIVE BREAST CANCER


