



Triple-negative breast cancer incidence in the United States: ecological correlations with area-level sociodemographics, healthcare, and health behaviors

Jennifer L. Moss^{1,2} · Zaria Tatalovich³ · Li Zhu³ · Camille Morgan⁴ · Kathleen A. Cronin³

Received: 10 January 2020 / Accepted: 2 July 2020 / Published online: 15 July 2020

© This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2020

Abstract

Purpose Triple-negative breast cancer (TNBC) is an aggressive subtype of breast cancer, more commonly diagnosed among black women than other subgroups. TNBC varies geographically, but little is known about area-level characteristics associated with elevated incidence.

Methods We generated 2011–2013 age-adjusted TNBC incidence rates for state economic areas (SEAs) in 43 states using data from North American Association of Central Cancer Registries. For cases missing data on molecular markers, we imputed TNBC status using cross-marginal proportions. We linked these data to SEA covariates from national sources. Using linear ecological regression, we examined correlates of TNBC incidence rates for the overall population and for age (< 50 years or 50+ years)- or race (white or black)-specific subgroups.

Results The mean annual incidence of TNBC across SEAs was 13.7 per 100,000 women (range = 4.5–26.3), with especially high and variable rates among African American women (mean = 20.5, range 0.0–155.1). TNBC incidence was highest in South Atlantic and East South Central Census Divisions and lowest in Mountain Division. Overall TNBC incidence was associated with SEA sociodemographics (e.g., percent of females age 45+ who are non-Hispanic black: coefficient estimate [est.] = 1.62), healthcare characteristics (e.g., percent of population without health insurance: est. = -0.52), and health behaviors (e.g., prevalence of obesity among women: est. = 0.72) (all $p < 0.05$). Other variables related to TNBC incidence included density of obstetrician/gynecologists and prevalence of smoking.

Conclusion TNBC incidence varied across SEAs in the U.S., particularly for African American women. Identifying areas with elevated TNBC incidence can facilitate research and interventions on area- and individual-level correlates of TNBC.

Keywords Triple-negative breast cancer · Geographic disparities · Sociodemographics · Healthcare · Health behaviors

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s12282-020-01132-w>) contains supplementary material, which is available to authorized users.

✉ Jennifer L. Moss
jmoss1@pennstatehealth.psu.edu

¹ Cancer Prevention Fellowship Program, National Cancer Institute, Bethesda, MD 20892, USA

² Department of Family and Community Medicine, Penn State College of Medicine, 134 Sipe Ave., #205, MC HS72, PO Box 850, Hershey, PA 17033, USA

³ Surveillance Research Program, National Cancer Institute, Bethesda, MD 20892, USA

⁴ Center for Global Health, National Cancer Institute, Bethesda, MD 20892, USA

Introduction

Breast cancer is the most commonly diagnosed cancer among women in the United States, with more than 250,000 new cases each year [1]. Molecular profiling of breast tumors provides insight into treatment and prognosis for breast cancer subtypes [2, 3]. The most common tumor subtype (73% of breast cancers) is characterized by positive estrogen receptors (ER) or progesterone receptors (PR), and negative human epidermal growth factor type 2 receptors (HER2) [2]; this subtype is associated with the most promising prognoses [4]. In contrast, the tumor subtype that is negative for ER, PR, and HER2 (i.e., triple-negative breast cancer, or TNBC) makes up 12% of breast cancers [2] and has the poorest prognosis [3, 4].

TNBC differs from other breast cancer subtypes in many ways. Compared to other breast cancer subtypes, TNBC is more commonly diagnosed among women who are African American [5–7], premenopausal [3, 5, 8], and low socioeconomic status (SES) [3] than other types of breast cancer. TNBC is positively associated with reproductive factors including early menarche [5, 6, 9] and high parity [5, 8, 9], but may be negatively associated with breastfeeding [5, 8, 9]. Another risk factor is obesity [5, 9–11], although the full nature of this relationship is not clear. Mammography screening is not closely linked to detection of TNBC [12], partly because many women develop TNBC before the age of routine screening [13] and partly due to fast-growing interval cancers [14]. In 2015, the Annual Report to the Nation on the Status of Cancer indicated that TNBC incidence varies geographically more than do other subtypes, with especially high rates in the Southeast [2].

Geographic variation in TNBC is likely due to a combination of individual-level factors (including those described above) and area-level factors. For example, high TNBC incidence in the Southeast could be due to the higher incidence of TNBC among African American women [5–7], who disproportionately live in Southeastern states. Other, more contextual explanations are also possible. For example, a sociological perspective could suggest that young women living in low-SES areas are exposed to more female-headed, unmarried households, resulting in earlier initiation of child-bearing [15]; younger age at first birth is, in turn, associated with increased risk of TNBC [16]. Ecological analyses of area-level correlates of TNBC incidence can examine these relationships to help prioritize future research inquiries about the etiology of TNBC and why it varies so markedly by geography.

In the current paper, we describe geographic patterns in TNBC incidence rates and identify ecological correlates of TNBC. We contrast these findings with patterns of (1) ER-positive/HER2-negative breast cancer incidence and (2) total breast cancer incidence to highlight differences across tumor subtypes. These findings can inform future research studies and public health interventions targeting TNBC.

Materials and methods

Data sources and variables

Dependent variable: age-adjusted breast cancer incidence rates

We calculated age-adjusted incidence rates of TNBC for State Economic Areas (SEAs) using data from the North American Association of Central Cancer Registries (NAACCR)'s Cancer in North American (CiNA)

restricted data files (<https://www.naacr.org/cina-in-seers-tat/>). NAACCR collects data from cancer registries for all 50 states and Washington, D.C. [17] (hereafter referred to collectively as “states”). Data are available at the county level, but due to concerns about patient confidentiality and data scarcity (72% of counties had < 16 cases of TNBC during the study period), we linked counties to the SEA level. SEAs are sets of neighboring counties within a state that have similar economic characteristics [18].

Currently, breast cancer records in CiNA include tumor data on molecular markers (i.e., ER, PR, and HER2). However, NAACCR only began requiring registries to include HER2 status in 2010, and the first year of data is considered incomplete. Thus, we calculated age-adjusted TNBC incidence rates (defined as breast cancers that were negative for ER, PR, and HER2 [2]) for 2011–2013, expressed as cases per 100,000 women per year. We calculated these rates for the overall female population, and then for subgroups defined by race (White or Black) or age (< 50 years or 50 + years).

For breast cancer cases that were missing data on molecular markers (7%, primarily missing HER2 status), we imputed TNBC status. We calculated the marginal proportions of TNBC versus other subtypes of breast cancer among known cases, and applied those proportions to assign unknown cases to TNBC versus other subtypes of breast cancer [19]. This procedure has demonstrated adequate equivalence to other methods of imputation [20]. For analyses stratified by age or race categories, we repeated this imputation procedure for each subgroup separately.

As a comparison outcome, we also estimated SEA incidence rates of ER + /HER2– breast cancers and of breast cancers and for 2011–2013.

Independent variables: SEA characteristics

We gathered data on area-level characteristics that have previously demonstrated associations with TNBC either at the individual or area level (for a full list of variables, see Supplementary Table S1). For each characteristic, data were gathered for counties but aggregated to SEAs. Independent variables were standardized to a mean of 0 and standard deviation of 1 prior to inclusion in regression models (described below).

SEA sociodemographic characteristics came from the American Community Survey (ACS) 2010–2014 five-year estimates, generated by the U.S. Census Bureau [21]. These variables include measures of the population composition by sex, age, and race/ethnicity (e.g., percent of females age 45 + who are non-Hispanic black) and SES indicators (e.g., percent of population living below 150% of the federal poverty line).

SEA healthcare access characteristics came from the ACS and the 2011–2013 Area Health Resource Files, compiled by the Health Resources and Services Administration [22]. These variables include measures of density of healthcare resources (e.g., number of primary care providers per 1000 population) and health insurance coverage (e.g., percent of population without health insurance).

Finally, SEA health behavior characteristics came from national health surveys. These variables include prevalence of obesity among females (defined as having a body mass index of 30 or greater, from the Behavioral Risk Factor Surveillance System [23]) and prevalence of recent mammography screening (defined as within the last 2 years among women ages 40+, from the National Cancer Institute's Small Area Estimates [24]).

Statistical analysis

First, we conducted descriptive analysis by examining the distribution of the independent variables and age-adjusted TNBC incidence rates (using imputed data) across SEAs. As part of this analysis, we created a choropleth map depicting TNBC incidence across SEAs. The choropleth maps indicate TNBC incidence across SEAs, with classes determined using the Jenks natural breaks classification method [25] to maximize variation between classes and to minimize it within classes. The choropleth map for the overall distribution of rates for the total population used one classification scheme. The choropleth map for the distribution of rates by age and by race used different classification schemes to facilitate comparisons across groups. As appropriate, SEAs with zero observed cases of TNBC, SEAs not included in the dataset, and SEAs with suppressed values (i.e., with fewer than 16 cases of TNBC [26]) are also indicated.

Next, we conducted ecological linear regression analyses examining the associations between the independent variables and age-adjusted TNBC incidence rates. Bivariate models regressed TNBC incidence (separately for the total population and for age- or race-specific subgroups) on each independent variable. To create a multivariable ecological linear regression model, we conducted backwards selection

for the overall TNBC incidence rates, eliminating independent variables until all remaining independent variables were associated ($p < 0.10$) with TNBC [27]. We used this set of independent variables to run multivariable linear regressions for TNBC across age- or race-specific subgroups. We noted the adjusted R^2 value to indicate how well these independent variables explain the variation in cancer incidence rates. Then we conducted a sensitivity analysis by repeating the multivariable linear regression analysis using TNBC incidence rates based only on the observed cases of TNBC (i.e., excluding the imputed cases).

Finally, as a comparison analysis, we conducted multivariable linear regression analysis for ER+/HER2- breast cancers and for all breast cancers (for the overall population only).

The current study analysis was approved by the NAACCR Research Application Review Workgroup and the NAACCR Institutional Review Board on Human Subjects. Forty-seven NAACCR registries (covering 42 states, Washington, D.C., and five metro areas) consented to participate in the current analysis; participating registries covered 2526 (of 3142) counties and 415 (of 508) SEAs. (States that did not have any data included in the present analysis were Florida, Illinois, Kansas, Maryland, Minnesota, Missouri, Nevada, and Vermont.) Analyses were conducted using SAS version 9.4 (Cary, NC), and maps were generated using ArcGIS 10.6 (ESRI, Inc., Redlands, CA). All statistical tests used a two-sided p value of 0.05.

Results

During 2011–2013, 595,789 cases of breast cancer were diagnosed in the included registries. Of these cases, an estimated 67,903 cases were TNBC (including imputed cases; 11.4% of all breast cancers). Across SEAs, the age-adjusted TNBC incidence rate for the overall population was 13.7 cases per 100,000 women per year (standard error [SE]=0.2), ranging from 4.5 to 26.3 (Table 1). As expected, TNBC incidence varied across age- or race-specific subgroups, with higher rates observed among older women

Table 1 Distribution of age-adjusted breast cancer incidence rates per 100,000 women per year across state economic areas ($n = 415$) in the United States, 2011–2013

	Triple-negative breast cancer			ER+/HER2- breast cancer			All breast cancer		
	Mean	SE	Range	Mean	SE	Range	Mean	SE	Range
All women	13.7	0.2	(4.5–26.3)	75.9	0.7	(21.7–115.2)	120.3	0.7	(65.4–161.6)
< 50 years	6.3	0.1	(1.0–22.4)	23.0	0.3	(8.3–38.9)	41.1	0.3	(19.8–62.2)
50+ years	33.1	0.4	(9.2–66.0)	214.5	2.0	(49.6–331.1)	327.7	1.8	(184.7–438.3)
White women	12.6	0.1	(4.6–26.9)	77.8	0.8	(22.6–118.2)	120.6	0.7	(65.8–162.4)
Black women	20.5	0.9	(0.0–155.1)	58.7	2.2	(0.0–627.2)	108.3	2.7	(0.0–627.2)

ER estrogen receptor, HER2 human epidermal growth factor type 2 receptors, SE standard error

(33.1 per 100,000; SE = 0.4; range across SEAs: 9.2–66.0) and black women (20.5 per 100,000; SE = 0.9; range across SEAs: 0.0 to 155.1). For the whole population, TNBC incidence rates were highest in the South Atlantic and East South Central Census Divisions and lowest in the Mountain Division (Fig. 1). For each age- (Fig. 2) and race- (Fig. 3) specific subgroup, the highest observed rates for that group generally clustered in the South Atlantic and East South Central areas, as well.

Bivariate associations with TNBC

In bivariate analysis, overall TNBC incidence (among women of all ages and all races) was higher in SEAs with more black women, female-headed households, and residents living in poverty; with greater densities of obstetrician/gynecologists (OBGYNs) and oncologists; and with higher prevalence of obesity, recent mammography, and smoking (all $p < 0.05$) (Table 2). In contrast, overall TNBC incidence was lower in SEAs with more non-literacy and with greater densities of mammography facilities (both $p < 0.05$). For example, for each one-unit increase in the concentration of black women over age 45, there were 2.04 additional cases of TNBC per 100,000 women per year ($p < 0.001$), and for each one-unit increase in non-literacy, there were 0.56 fewer cases of TNBC ($p < 0.01$).

These patterns of associations were relatively similar across age- or race-specific subgroups. However, the magnitude of these associations varied; for example, among white women, relatively small differences in TNBC incidence were observed by the concentration of black women over age 45 in an SEA (estimate (est.) = 0.30, $p < 0.05$), but this association was much stronger for TNBC incidence among black women (est. = 3.38, $p < 0.001$) and older women (est. = 4.57,

$p < 0.001$). In general, relationships between the independent variables and TNBC incidence were stronger among older women compared to younger women, and among black women compared to white women.

Multivariable associations with TNBC

In multivariable analysis, we retained seven independent variables that captured sociodemographic, healthcare, and health behavior characteristics across SEAs (Table 3). Overall TNBC incidence was higher in SEAs with more black women (est. = 1.62), greater densities of OBGYNs (est. = 0.40), and higher prevalence of obesity (est. = 0.72) and smoking (est. = 0.63) (all $p < 0.05$) (adjusted $R^2 = 0.36$). In contrast, overall TNBC incidence was lower in SEAs with more working class residents (est. = -0.55) and more residents without health insurance (est. = -0.52) (both $p < 0.05$).

Generally, these patterns of associations were similar across age- or race-specific subgroups, although many coefficient estimates lost statistical significance and the adjusted R^2 values were smaller in the subgroup analyses. Thus, none of the independent variables were associated with TNBC incidence for all of the age- or race-specific subgroups. Again, the magnitude of the associations varied. Among white women, TNBC incidence was not associated with the concentration of black women (est. = -0.05, $p < 0.05$), but this association was large and statistically significant among older women (est. = 3.44, $p < 0.001$). In general, relationships between the independent variables and TNBC incidence were stronger among older women compared to younger women, and among black women compared to white women.

Fig. 1 Age-adjusted incidence rate of triple-negative breast cancer, per 100,000 women, in state economic areas ($n = 415$) in the United States, 2011–2013. Values are suppressed for areas with < 16 cases

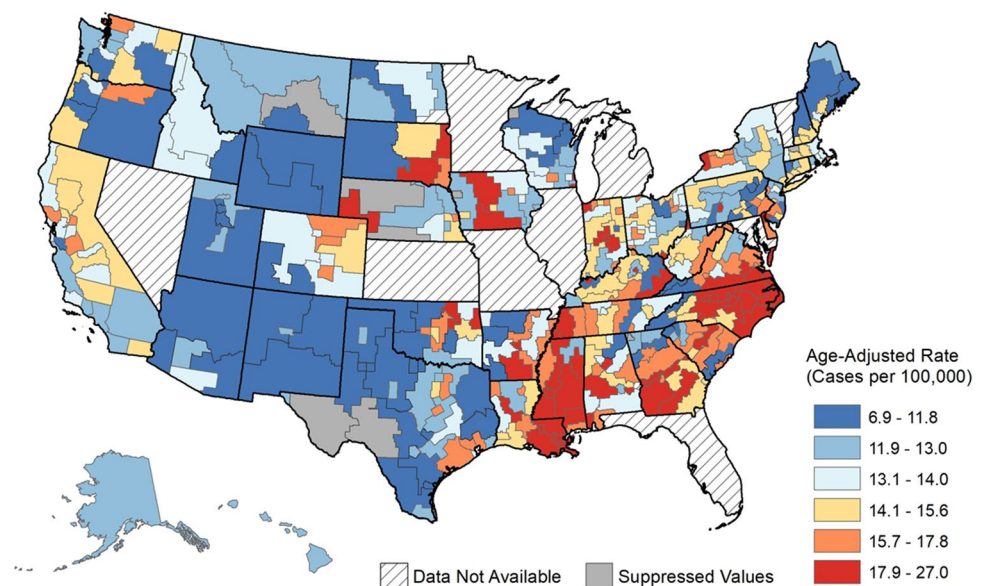
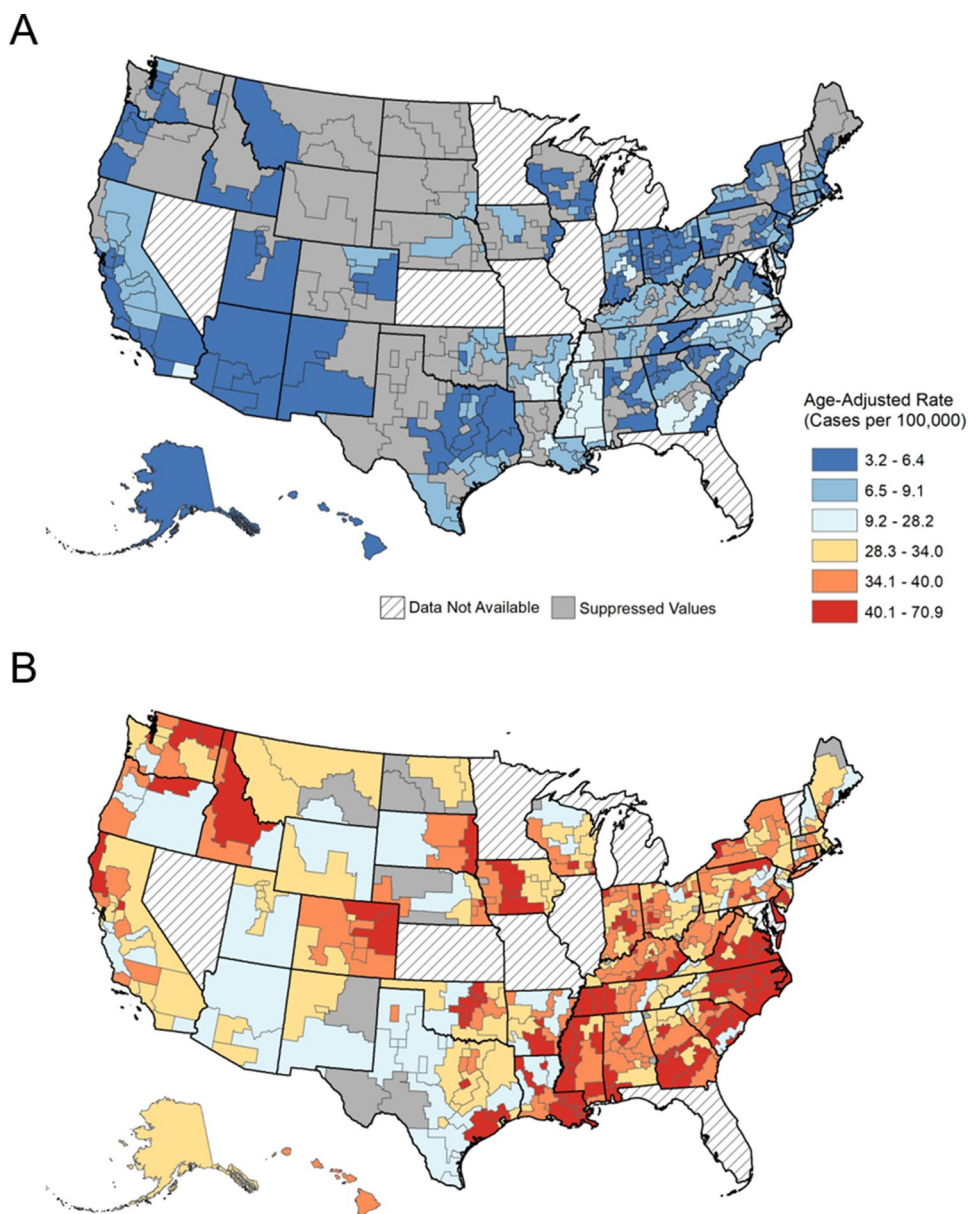


Fig. 2 Age-adjusted incidence rate of triple-negative breast cancer, per 100,000 women, for **a** women <50 years and **b** women 50+ years in state economic areas ($n=415$) in the United States, 2011–2013. Values are suppressed for areas with < 16 cases



Sensitivity analysis

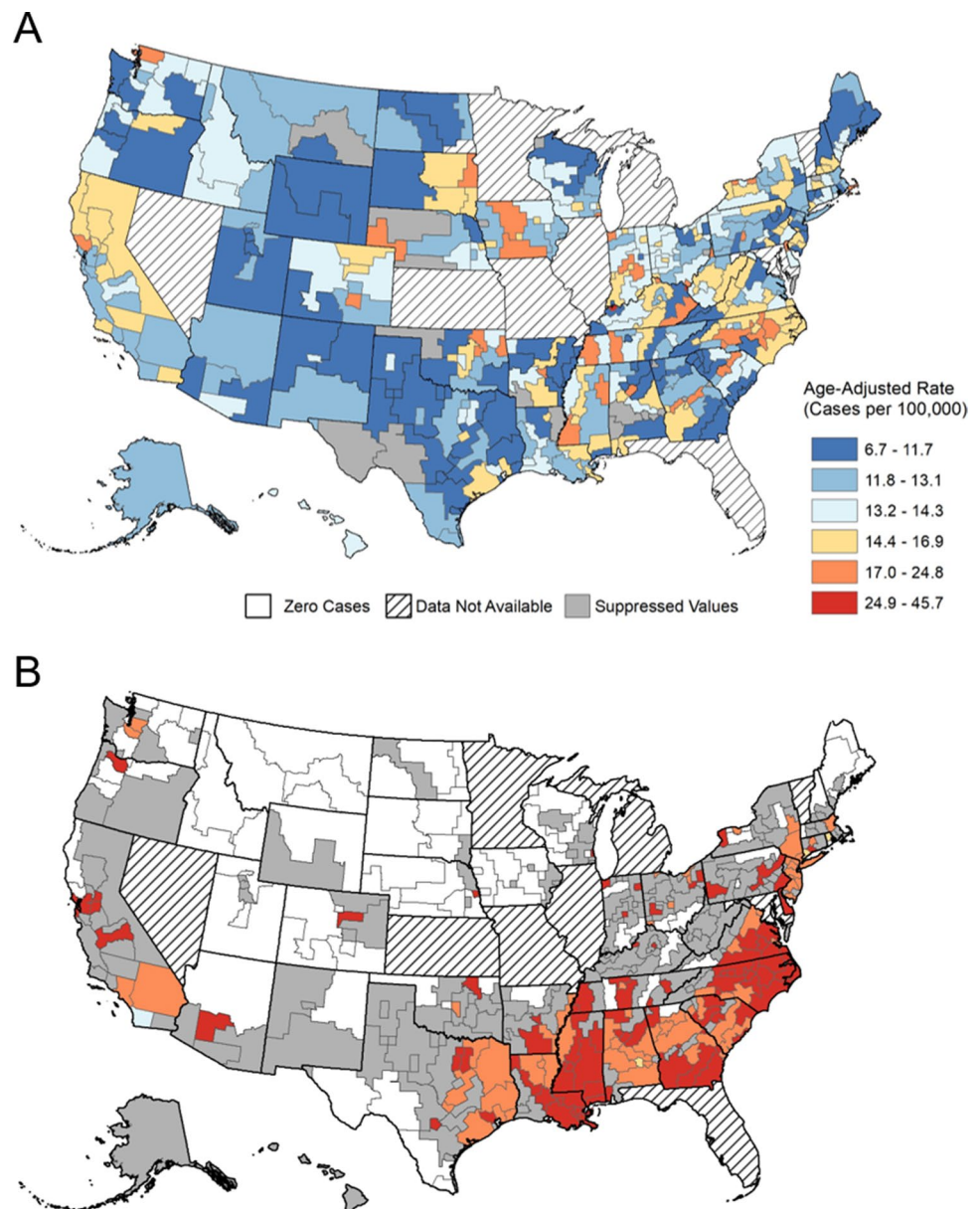
When we repeated the multivariable models using incidence rates based on only the observed cases of TNBC, we found very similar associations with the SEA characteristics as in the main analysis (Supplementary Table S2). All of the model estimates maintained the same direction and similar magnitude, with only a few differences in statistical significance compared to the models using the imputed data (Table 3).

Comparison of associations for TNBC versus other breast cancers

Bivariate (Table 2) and multivariable (Table 3) models examining TNBC for all ages and all races were

relatively similar to the results of models examining ER+/HER2– breast cancer and all breast cancers across SEAs (Supplementary Table S3). A major exception concerns the associations between breast cancer and health behaviors. In multivariable models, TNBC incidence was positively associated with prevalence of obesity and smoking; however, these behaviors were negatively associated with incidence of ER+/HER2– and all breast cancers (although these were only statistically significant for ER+/HER2– cancers). Other variables that demonstrated distinct associations with incidence for the different types of breast cancer are population without health insurance (est. = -0.52 , $p < 0.05$ for TNBC; est. = -6.63 , $p < 0.001$ for ER+/HER2–; est. = -5.07 , $p < 0.001$ for all breast cancer) and OBGYN density (est. = 0.40 , $p < 0.05$ for TNBC; est. = 2.19 , $p < 0.01$

Fig. 3 Age-adjusted incidence rate of triple-negative breast cancer, per 100,000 women, for **a** white women and **b** black women in state economic areas ($n=415$) in the United States, 2011–2013. Values are suppressed for areas with <16 cases



for ER+ /HER2−; est. = 2.50, $p < 0.001$ for all breast cancer).

Discussion

In this analysis of more than half a million cases of breast cancer across 42 states and Washington, D.C., we found striking geographic variation in the incidence of TNBC. The highest rates of TNBC incidence were observed primarily in the southeastern regions, and the lowest rates were generally in the western parts of the country. For the entire population of women, TNBC incidence was 13.7 cases per 100,000 women (SE = 0.2; range = 4.5–26.3) across SEAs, but rates were especially high for older women (33.1, SE = 0.4) and

black women (20.5, SE = 0.9). In multivariable analysis, TNBC incidence was associated with several SEA characteristics, particularly racial/ethnic composition, percent uninsured, and prevalence of smoking and obesity. This pattern of finding was relatively consistent across subgroups of women and in our sensitivity and comparison analyses.

Among sociodemographics of SEAs, we found associations between TNBC incidence and the percent of (1) women who were black and age 45+ and (2) people in the labor force in working class occupations. TNBC incidence was higher in areas that had higher densities of older black women in models examining incidence among all women and in age-specific subgroups. This finding is expected given the elevated rates in older women and black women observed in the current analysis and in the extant literature

Table 2 Bivariate associations between independent variables and triple-negative breast cancer incidence rates across state economic areas ($n=415$) in the United States, 2011–2013

	All women est.	<50 years est.	50+ years est.	White est.	Black est.
Percent female (45+)	-0.04	0.07	-0.34	0.03	-0.50
Percent black	2.04***	1.07***	4.57***	0.30*	3.38***
Urbanicity	0.31	0.09	0.90*	0.21	1.11
Kids per household	0.14	0.19	0.00	-0.25	-0.29
Female-headed households	1.65***	0.96***	3.47***	0.27	2.77**
Females with HS degree	-0.18	-0.33**	0.23	0.18	-1.11
English non-literacy	-0.56**	-0.26*	-1.34**	-0.25	-0.84
Poverty	0.62***	0.50***	0.93*	0.00	0.64
Working class	-0.32	0.01	-1.19**	-0.26	-0.55
PCP density	0.33	0.01	1.18**	0.34*	-0.54
OBGYN density	1.02***	0.41**	2.65***	0.42**	1.64
Oncologist density	0.73***	0.23	2.05***	0.41**	1.18
Mammography facility density	-0.35*	-0.11	-0.99*	-0.15	-2.84**
Population without health insurance	-0.06	0.11	-0.49	-0.40**	0.42
Population with public health insurance	0.33	0.38**	0.21	-0.03	0.38
Obesity prevalence (females)	1.27***	0.78***	2.54***	0.28*	2.82**
Recent mammography prevalence	0.51**	0.11	1.57***	0.29*	0.64
Current smokers prevalence	0.36*	0.41**	0.23	0.15	1.59

Linear regression models regress the respective TNBC incidence rate on the independent variable *est* coefficient estimate, *HS* high school, *PCP* primary care provider, *OBGYN* obstetrician/gynecologist * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 3 Multivariable associations between independent variables and triple-negative breast cancer incidence rates across state economic areas ($n=415$) in the United States, 2011–2013

	All women est.	<50 years est.	50+ years est.	White est.	Black est.
Percent black	1.62***	0.92***	3.44***	-0.05	1.81
English non-literacy	0.41	0.51**	0.14	0.41	2.48
Working class	-0.55*	-0.38	-0.98	-0.41	-3.72*
OBGYN density	0.40*	0.13	1.12*	0.38*	0.44
Population without health insurance	-0.52*	-0.26	-1.21*	-0.56**	-0.44
Obesity prevalence (females)	0.72**	0.31	1.78*	0.65*	2.77
Current smokers prevalence	0.63*	0.80***	0.20	0.44	3.87*
R^2	0.36	0.21	0.30	0.06	0.04

Linear regression models regress the respective TNBC incidence rate on the independent variables. R^2 indicates the proportion of variation in the triple-negative breast cancer incidence rates explained by the combined contribution of the independent variables

Est. coefficient estimate, *OBGYN* obstetrician/gynecologist * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

[2, 5, 7]. Interestingly, these associations in the multivariable models varied considerably in magnitude, with an especially strong relationship for older women (of any race); this association suggests that areas with high concentrations of older black women are especially vulnerable to TNBC, but it is unclear why. The association with working class, however, is more difficult to interpret. Working class employment is often observed to be a risk factor for cancer outcomes [28], but in the current study, it was protective: TNBC incidence rates were lower in SEAs with

more working class workers (for all women and for Black women). Several potential explanations exist, including the fact that TNBC has not been linked to occupational exposures (as in the relationship between exposure to agricultural pesticides and non-Hodgkin lymphoma [29]) and living in an area with more working class workers could protect women from the more deleterious effects of living in an area with higher unemployment [30]. Additional research is needed to clarify this association, especially for black women.

Among SEA healthcare characteristics, we found associations between TNBC incidence and (1) the number of OBGYNs per 1000 people and (2) the percent of people without health insurance. For both variables, TNBC incidence was greater in areas that were marked by increased availability and accessibility of healthcare resources (i.e., higher densities of OBGYNs; lower percentages of uninsured). (Note that, in addition, in bivariate analysis, TNBC incidence was positively associated with density of other types of healthcare providers, although these relationships did not maintain statistical significance in multivariable analysis.) These findings are counterintuitive, since research studies often find that cancer outcomes are worse in areas with lower geographic access to healthcare [31]. However, healthcare services may be limited in their potential for lowering TNBC incidence, given that there are few prevention strategies and precursor markers for this type of breast cancer [32]. Instead, these healthcare characteristics may reflect other dimensions of area-level SES (beyond working class, noted above); indeed, SEA-level OBGYN density and percent uninsured were highly correlated with percent working class ($r = -0.62$ and $r = 0.39$, respectively; both $p < 0.05$; data not shown). One possibility is that these healthcare characteristics reflect constructs such as urbanicity, population density, and/or access to care; although urbanicity was not associated with TNBC incidence in bivariate analyses, it is notoriously difficult to measure aspects of urbanicity relevant to healthcare outcomes [33]. The observed multivariable associations between healthcare characteristics and TNBC incidence were limited to the subgroups of older women and white women, potentially indicating that the healthcare environment is less germane to TNBC risk among other women.

Among SEA health behaviors, we found associations between TNBC incidence and the prevalence of (1) obesity among women and (2) smoking among all people. TNBC incidence rates (for all women, older women, and white women) were higher in areas with greater prevalence of obesity, and TNBC incidence rates (for all women, younger women, and black women) were higher in areas with greater prevalence of smoking. Thus, the two behaviors correlate with TNBC for different age- or race-specific subgroups; understanding the potential differential risk associated with these behaviors for TNBC among different subgroups is an area for future research. At the individual level, obesity is positively associated with risk of TNBC [5, 9, 10], but several previous studies have demonstrated a null relationship between smoking and TNBC risk [9, 10, 34–36]. Areas that currently have relatively high smoking rates, such as states in the Southeast [37], may have other risks for TNBC, such as higher densities of black residents [38], lower SES [39], higher parity [40], and less generous welfare and Medicaid policies [41]. Understanding whether aggregated measures

of health behaviors (e.g., obesity, smoking) function as proxies for individual health behaviors or as confounders for other, more important contextual factors (such as social welfare policies [41]) is an important next step for probing the geographic variation in TNBC.

Taken together, these ecological findings indicate that TNBC incidence rates are highest in regions marked by sociodemographic, healthcare, and health behavior challenges. These results complement individual-level analyses that highlight risk factors including race/ethnicity, SES, age, genetics, and reproductive factors [2, 3, 5, 9, 10]. While we had a tentative hypothesis that higher area-level SES (especially around educational attainment, employment, and family structure among women) would correlate with lower TNBC incidence (because young women would have role models to encourage them to delay early childbearing [16]), we did not find evidence for this relationship. More theoretical development is needed to explicate the observed ecological relationships, as well as to situate these findings within the context of individual-level risk factors. Additional research is needed to examine these ecological relationships in different settings, e.g., at different geographic units or in different countries. It is possible, for example, that in other countries with different healthcare systems, the proportion of the population without health insurance may not be meaningfully related to TNBC. In addition, this study (and other studies) have indicated that TNBC is more common among black women and in black communities; in the U.S., ‘black’ race includes individuals with incredibly different origins and ancestries. The relationship between TNBC and race/ethnicity likely varies in countries outside the U.S. that have different (1) social structures and (2) populations with racial/ethnic profiles.

From a methodological standpoint, this study used several best practices for dealing with sparse data and maintaining patient confidentiality. First, we preserved as much sample size as possible by imputing triple-negative status for breast cancer cases with missing data on hormonal markers. Next, we aggregated data to a relatively large geographic area, SEA, to accumulate enough cases in each unit to allow for stable estimation of age-adjusted rates (including rates stratified by age and race). Finally, we used geographic analysis methods to examine ecological associations with TNBC incidence; although individual-level and ecological analyses are complementary, in many cases, individual-level data are not available. Other researchers interested in the epidemiology of rare cancers and/or subtypes may benefit from using similar approaches.

In terms of study strengths, we used a large, population-based dataset covering cancer patients throughout most of the United States. The data from these cancer registries provide near-100% coverage of cancer cases in their respective catchment areas [17], increasing our confidence

in the validity of the findings. In addition, we used independent variables capturing area-level variation in several domains (i.e., sociodemographics, healthcare, and health behavior) relevant to TNBC incidence. Finally, we found consistency across the findings from the main analysis, the sensitivity analysis, and the comparison analysis, with some important (and expected) exceptions. In terms of study limitations, our data only included 3 years of TNBC incidence, since NAACCR only began requiring registries to report HER2 status in 2010; additional years of data will increase the stability of the observed TNBC incidence rates. A related limitation is that 7% of breast cancer cases had incomplete data on hormonal markers; however, we employed a previously-tested imputation technique [19] to address this limitation, and findings were relatively similar when we restricted the analysis to only cases with complete hormonal information (although alternative imputation approaches may find slightly different estimates of rates and standard errors [20]). In addition, we had missing data because registries from eight states declined to participate in this study; some bias in the observed associations could result from incomplete geographic coverage. Another limitation is that, due to concerns about patient confidentiality and statistical stability, data were aggregated to a relatively large geographic area (i.e., state economic areas); using alternative levels of spatial aggregation may reveal different relationships [42]. Some risk factors for TNBC, such as BRCA1 mutations [43, 44], are not available at the ecological level; multilevel analyses that incorporate individual- and area-level factors are an important next step to overcome this limitation. Multilevel analyses which include individual-level risk factors would likely also explain much more of the variation in TNBC incidence than the ecological models in the current study. Finally, we used backwards stepwise regression to develop a multivariable model to use with all the outcome variables. This technique is limited [27], but we chose to use this approach to ensure comparability of models analyzing TNBC for all the subgroups.

In conclusion, we found great geographic variation in TNBC incidence in an analysis of more than half a million cases of breast cancer from cancer registries in 43 states, with especially high rates among older women and black women. TNBC incidence rates for SEAs correlated with sociodemographics (e.g., racial/ethnic composition), healthcare characteristics (e.g., insurance coverage), and health behaviors (e.g., prevalence of obesity). These findings have important implications for additional research to (1) create theoretically informed hypotheses about how these ecological variables influence individual-level risk of TNBC and (2) integrate ecological and individual-level variables into a more comprehensive analysis of TNBC risk.

Acknowledgements These data are based on the NAACCR December 2016 data submission. Support for cancer registries is provided by the state, province or territory in which the registry is located. In the U.S., registries also participate in the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program or the Centers for Disease Control and Prevention's National Program of Cancer Registries (NPCR) or both. In Canada, all registries submit data to the Canadian Cancer Registry maintained by Statistics Canada. The opinions expressed in this article are the authors' own and do not reflect the view of the National Institutes of Health, the Department of Health and Human Services, or the United States government.

Author contributions JLM obtained, analyzed, and interpreted the data and drafted the manuscript. ZT assisted with data analysis and interpretation, and revised the manuscript. LZ assisted with data analysis and interpretation, and revised the manuscript. CM contributed to data cleaning, analysis, and interpretation, and revised the manuscript. KAC provided study oversight and assisted with data interpretation, and revised the manuscript. All authors read and approved the final manuscript.

Funding The authors completed this work as part of active duty at the National Cancer Institute. No additional funding to support this project was received.

Compliance with ethical standards

Conflict of interest The authors have no potential conflicts of interest to disclose.

References

1. National Cancer Institute. Cancer stat facts: Female breast cancer. 2018; <https://seer.cancer.gov/statfacts/html/breast.html> (2018).
2. Kohler BA, Sherman RL, Howlader N, et al. Annual report to the nation on the status of cancer, 1975–2011, featuring incidence of breast cancer subtypes by race/ethnicity, poverty, and state. *J Natl Cancer Inst.* 2015;107(6):dju048.
3. Bauer KR, Brown M, Cress RD, Parise CA, Caggiano V. Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype. *Cancer.* 2007;109(9):1721–8.
4. Carey LA, Perou CM, Livasy CA, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA.* 2006;295(21):2492–502.
5. Dawood S. Triple-negative breast cancer. *Drugs.* 2010;70(17):2247–58.
6. Yang XR, Sherman ME, Rimm DL, et al. Differences in risk factors for breast cancer molecular subtypes in a population-based study. *Cancer Epidemiol Prev Biomarkers.* 2007;16(3):439–43.
7. Howlader N, Altekruse SF, Li CI, et al. US incidence of breast cancer subtypes defined by joint hormone receptor and HER2 status. *J Natl Cancer Inst.* 2014;106(5):dju055.
8. Chollet-Hinton L, Anders CK, Tse CK, et al. Breast cancer biologic and etiologic heterogeneity by young age and menopausal status in the Carolina Breast Cancer Study: a case-control study. *Breast Cancer Res.* 2016;18(1):79.
9. Millikan RC, Newman B, Tse CK, et al. Epidemiology of basal-like breast cancer. *Breast Cancer Res Treat.* 2008;109(1):123–39.

10. Turkoz FP, Solak M, Petekkaya I, et al. Association between common risk factors and molecular subtypes in breast cancer patients. *Breast*. 2013;22(3):344–50.
11. Bandera EV, Chandran U, Hong C-C, et al. Obesity, body fat distribution, and risk of breast cancer subtypes in African American women participating in the AMBER Consortium. *Breast Cancer Res Treat*. 2015;150(3):655–66.
12. Wheeler SB, Reeder-Hayes KE, Carey LA. Disparities in breast cancer treatment and outcomes: biological, social, and health system determinants and opportunities for research. *Oncologist*. 2013;18(9):986–93.
13. U. S. Preventive Services Task Force. Final update summary: Breast cancer: Screening. (2016).
14. Collett K, Stefansson IM, Eide J, et al. A basal epithelial phenotype is more frequent in interval breast cancers compared with screen detected tumors. *Cancer Epidemiol Prev Biomarkers*. 2005;14(5):1108–12.
15. South SJ, Baumer EP. Deciphering community and race effects on adolescent premarital childbearing. *Soc Forces*. 2000;78(4):1379–407.
16. Anderson KN, Schwab RB, Martinez ME. Reproductive risk factors and breast cancer subtypes: a review of the literature. *Breast Cancer Res Treat*. 2014;144(1):1–10.
17. North American Association of Central Cancer Registries. Data standards & data dictionary, volume II. 2018; <https://www.naaccr.org/data-standards-data-dictionary/> (2018).
18. U.S. Census Bureau. Geographic areas reference manual. 2015; <https://www.census.gov/geo/reference/garm.html> (2018)
19. Anderson WF, Katki HA, Rosenberg PS. Incidence of breast cancer in the United States: current and future trends. *J Natl Cancer Inst*. 2011;103(18):1397–402.
20. Howlader N, Noone A-M, Yu M, Cronin KA. Use of imputed population-based cancer registry data as a method of accounting for missing information: application to estrogen receptor status for breast cancer. *Am J Epidemiol*. 2012;176(4):347–56.
21. U. S. Census Bureau. American Community Survey (ACS). 2018; <https://www.census.gov/programs-surveys/acs/>, (2018).
22. U. S. Department of Health and Human Services. Area Health Resources Files 2016; <https://data.hrsa.gov/topics/health-workforce/ahrf> (2018).
23. Centers for Disease Control and Prevention. Behavioral Risk Factor Surveillance System. 2014; <https://www.cdc.gov/brfss/> (2014).
24. National Cancer Institute. Small area estimates for cancer risk factors & screening behaviors. 2016; <https://sae.cancer.gov/>, (2018)
25. Brewer CA, Pickle L. Evaluation of methods for classifying epidemiological data on choropleth maps in series. *Ann Assoc Am Geogr*. 2002;92(4):662–81.
26. Richards TB, Berkowitz Z, Thomas CC, et al. Choropleth map design for cancer incidence, part 2. *Prev Chronic Dis*. 2010;7(1):A24.
27. Cohen J, Cohen P, West SG, Aiken LS. Applied multiple regression/correlation analysis for the behavioral sciences, vol. 3. London: Lawrence Erlbaum Associates; 2003.
28. Kogevas M, Porta M. Socioeconomic differences in cancer survival: a review of the evidence. *IARC Sci Publ*. 1997;138(138):177–206.
29. Schinasi L, Leon ME. Non-Hodgkin lymphoma and occupational exposure to agricultural pesticide chemical groups and active ingredients: a systematic review and meta-analysis. *Int J Environ Res Public Health*. 2014;11(4):4449–527.
30. Singh GK, Siahpush M. Inequalities in US life expectancy by area unemployment level, 1990–2010. *Scientifica*. 2016;2016:8290435.
31. Ambroggi M, Biasini C, Del Giovane C, Fornari F, Cavanna L. Distance as a barrier to cancer diagnosis and treatment: review of the literature. *Oncologist*. 2015;20(12):1378–85.
32. Schneider BP, Winer EP, Foulkes WD, et al. Triple-negative breast cancer: risk factors to potential targets. *Clin Cancer Res*. 2008;14(24):8010–8.
33. Meilleur A, Subramanian SV, Plascak JJ, Fisher JL, Paskett ED, Lamont EB. Rural residence and cancer outcomes in the United States: issues and challenges. *Cancer Epidemiol Biomark Prev*. 2013;22(10):1657–67.
34. Kabat GC, Kim M, Phipps AI, et al. Smoking and alcohol consumption in relation to risk of triple-negative breast cancer in a cohort of postmenopausal women. *Cancer Causes Control*. 2011;22(5):775–83.
35. Dolle JM, Daling JR, White E, et al. Risk factors for triple-negative breast cancer in women under the age of 45 years. *Cancer Epidemiol Biomarkers Prev*. 2009;18(4):1157–66.
36. Kawai M, Malone KE, Tang MT, Li CI. Active smoking and the risk of estrogen receptor-positive and triple-negative breast cancer among women ages 20 to 44 years. *Cancer*. 2014;120(7):1026–34.
37. Dwyer-Lindgren L, Mokdad AH, Srebotnjak T, Flaxman AD, Hansen GM, Murray CJ. Cigarette smoking prevalence in US counties: 1996–2012. *Popul Health Metrics*. 2014;12(1):5.
38. Datta GD, Subramanian SV, Colditz GA, Kawachi I, Palmer JR, Rosenberg L. Individual, neighborhood, and state-level predictors of smoking among US Black women: a multilevel analysis. *Soc Sci Med*. 2006;63(4):1034–44.
39. Guzman GG. Household income. *Am Commun Surv Br*. 2016;2017:1–7.
40. Mathews T, Hamilton BE. Total fertility rates by state and race and Hispanic origin: United States, 2017. *Natl Vital Stat Rep*. 2019;68(1):1–11.
41. Beckfield J, Bambra C. Shorter lives in stingier states: Social policy shortcomings help explain the US mortality disadvantage. *Soc Sci Med*. 1982;2016(171):30–8.
42. Krieger N, Chen JT, Waterman PD, Soobader MJ, Subramanian SV, Carson R. Geocoding and monitoring of US socioeconomic inequalities in mortality and cancer incidence: does the choice of area-based measure and geographic level matter?: the Public Health Disparities Geocoding Project. *Am J Epidemiol*. 2002;156(5):471–82.
43. Rakha EA, Reis-Filho JS, Ellis IO. Basal-like breast cancer: a critical review. *J Clin Oncol*. 2008;26(15):2568–81.
44. Reis-Filho JS, Tutt AN. Triple negative tumours: a critical review. *Histopathology*. 2008;52(1):108–18.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.