



# Integrating access to care and tumor patterns by race and age in the Carolina Breast Cancer Study, 2008–2013

Marc A. Emerson<sup>1</sup> · Yvonne M. Golightly<sup>1</sup> · Xianming Tan<sup>2,3</sup> · Allison E. Aiello<sup>1</sup> · Katherine E. Reeder-Hayes<sup>2</sup> · Andrew F. Olshan<sup>1,2</sup> · H. Shelton Earp<sup>2</sup> · Melissa A. Troester<sup>1</sup>

Received: 20 June 2019 / Accepted: 30 December 2019 / Published online: 16 January 2020  
© Springer Nature Switzerland AG 2020

## Abstract

**Purpose** Understanding breast cancer mortality disparities by race and age is complex due to disease heterogeneity, comorbid disease, and the range of factors influencing access to care. It is important to understand how these factors group together within patients.

**Methods** We compared socioeconomic status (SES) and comorbidity factors in the Carolina Breast Cancer Study Phase 3 (CBCS3, 2008–2013) to those for North Carolina using the 2010 Behavioral Risk Factor Surveillance Study. In addition, we used latent class analysis of CBCS3 data to identify covariate patterns by SES/comorbidities, barriers to care, and tumor characteristics and examined their associations with race and age using multinomial logistic regression.

**Results** Major SES and comorbidity patterns in CBCS3 participants were generally similar to patterns in the state. Latent classes were identified for SES/comorbidities, barriers to care, and tumor characteristics that varied by race and age. Compared to white women, black women had lower SES (odds ratio (OR) 6.3, 95% confidence interval (CI) 5.2, 7.8), more barriers to care (OR 5.6, 95% CI 3.9, 8.1) and several aggregated tumor aggressiveness features. Compared to older women, younger women had higher SES (OR 0.5, 95% CI 0.4, 0.6), more barriers to care (OR 2.1, 95% CI 1.6, 2.9) and aggregated tumor aggressiveness features.

**Conclusions** CBCS3 is representative of North Carolina on comparable factors. Patterns of access to care and tumor characteristics are intertwined with race and age, suggesting that interventions to address disparities will need to target both access and biology.

## Introduction

US population-based breast cancer mortality rates are 42% higher for black women compared to white women and the mortality disparity is present at all ages and diagnostic stages [1, 2]. Since the early 1980s, this mortality gap has

increased, partly reflecting disproportionate increases in survival among white women [1]. When considering age, black women age 45–54 have an absolute breast cancer mortality difference twice as high as white women the same age [3]. Understanding breast cancer disparities by race and age is complex, due to the biological heterogeneity of the disease, presence of comorbid conditions, and the multifaceted interaction of social, behavioral, and access to care and care quality factors [4, 5]. Disentangling the biological and access factors that drive breast cancer mortality is critical to developing effective interventions.

To address the knowledge gap of how race, age, social determinants, comorbid conditions and tumor characteristics interact, we used data from 2,998 women with invasive breast cancer in the population-based Carolina Breast Cancer Study Phase 3 (CBCS3) (2008–2013), which oversampled younger (< 50 age at diagnosis) women and black women. We used latent class analysis (LCA), a person-oriented approach to reduce highly dimensional data and

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s10552-019-01265-0>) contains supplementary material, which is available to authorized users.

---

✉ Melissa A. Troester  
troester@email.unc.edu

<sup>1</sup> Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

<sup>2</sup> Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

<sup>3</sup> Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

capture complex patterns that vary between people [6–9]. To demonstrate the generalizability of CBCS3 and better understand how well this cohort represented the source population, we also compared SES and access characteristics to the underlying population of North Carolina, using Behavioral Risk Factor Surveillance Study (BRFSS) data. North Carolina is a socio-geographically and ethnically diverse state with wide ranging rural areas as well as focal urban areas and corresponding differences in access to care, income and education. The resulting data identify key person-centered groupings of patients and important multivariable contributors to breast cancer mortality.

## Methods

### Data sources and study populations

CBCS3 is a population-based case-only study that was initiated to comprehensively evaluate the survivorship following invasive breast cancer diagnosis. All cases were identified within two months of diagnosis by rapid case ascertainment via the UNC Rapid Case Ascertainment Core in conjunction with North Carolina Central Cancer Registry. Younger (< 50 years in age) and black cases were oversampled by randomized recruitment so that half of the population was younger and half was black. All procedures performed in CBCS3 were in accordance with the ethical standards of the Institutional Review Board of the University of North Carolina at Chapel Hill. Informed consent was obtained from all participants. Eligibility criteria defined participants as female, between the ages of 20 and 74 at the time of diagnosis, and receiving a first, primary diagnosis of breast cancer between May 1, 2008 and October 21, 2013 with residence in the 44-county study area.

For comparisons of CBCS3 to the general North Carolina population, we examined the 2010 Behavioral Risk Factor Surveillance System (BRFSS), a state-based random telephone survey with a cross-sectional study design. To be eligible for inclusion, participants were age 18 and older, resided in households within any North Carolina county. BRFSS interviews are conducted monthly and data are collected and analyzed annually, and we utilized 2010 data, a midpoint time relative to CBCS3 study period. BRFSS data are weighted for the probability of telephone number selection, the number of adults in a household, and the number of phones in a household and are adjusted to reflect the demographic distribution of North Carolina's adult (age 18 and older) population. To be comparable to the CBCS3 study population, we restricted the BRFSS population to women who self-identified as either black or non-Hispanic white and were between the ages of 20 and 74 years in age. The lead author (MAE) created and signed a Data Use Agreement for

NC BRFSS for the analysis of publicly available NC BRFSS data. Sampling weights were applied to both data sources to match distributions in the original source population; in CBCS3 the source population was the 44-county study area and for NC BRFSS it was the entire state of North Carolina. All presented SES and comorbid factors were assessed in both CBCS3 and NC BRFSS. The categorization of these variables differed between datasets, however, we harmonized the CBCS3 and BRFSS for each categorical variable.

### SES and Comorbidities Factors: CBCS3 and NC BRFSS

For CBCS3, SES and comorbidity information were assessed by a baseline questionnaire and nurse-administered questionnaire on family history and body mass index (BMI) measurement within, on average, 5 months of diagnosis. For 2010 NC BRFSS, SES and comorbid information were assessed via a landline telephone survey. SES variables of interest included self-reported race (white vs. black), age (age at diagnosis in CBCS3; age at interview in BRFSS) (< 50 years of age vs. ≥ 50 years of age), marital status (married vs. not married), income (USD > \$50K, \$15K to \$50K and < \$15K), education (college degree or higher, some college, technical or business school, high school graduate/GED and 0–12 years, no but high school degree), current health insurance (yes vs. no) and rural address (yes vs. no).

Education, income and marital status had comparable categories between CBCS3 and BRFSS and were harmonized by categorization. For health insurance, CBCS3 participants were asked at baseline if they currently had health insurance coverage and the type of insurance (private health insurance purchased on their own or by husband or partner, private health insurance from their employer or workplace or that of their husband or partner, Medicaid, Medicare or other insurance that covered part of their medical bills). In BRFSS, participants were asked, “Do you have any kind of health care coverage, including health insurance, prepaid plans as HMOs, or government plans such a Medicare?”. A “yes” response was coded as current health care coverage. The variable on current health insurance was dichotomous: “yes vs. no”. For rurality, CBCS3 participants were asked about their community type since age 25. We then collapsed the categorization to include city (large city [population > 100K], suburb, and town or city with a population of < 10K, 10–50K and 50–100K) vs. rural (rural, non-farm, in the country and on a farm). In BRFSS, rural status was assessed based on metropolitan statistical areas (MSA), which are defined by the US Office of Management and Budget as a metropolitan area distinct from another metropolitan area. MSA codes included: in the center city of an MSA, outside the center city of an MSA but inside the county containing the center city, inside a suburban county of the MSA and not in an MSA. Rural status included the category of “not in an MSA”

while all other category we coded as “non-rural”. The final rural variable was coded as “yes vs. no”.

Comorbidity factors included diabetes (yes vs. no), heart disease (yes vs. no), smoking status (not current vs. current), and body mass index (BMI < 25 kg/m<sup>2</sup>, 25–30 kg/m<sup>2</sup> and 30 kg/m<sup>2</sup>). In CBCS3, diabetes and heart disease were determined by medical record abstraction as a comorbidity to breast cancer and were dichotomized as “yes vs. no”. In BRFSS, participants were asked, “have you ever been told by a doctor that you have diabetes?” If “Yes” and the respondent was female, the participant was asked, “Was this only when you were pregnant?” An affirmative response was coded as pre-diabetes or borderline if answered yes, which was less than 1% frequency. Participants were also asked if they had ever been told they had angina or coronary heart disease, an affirmative response was coded as “yes”. For smoking, CBCS3 participants were asked about their current tobacco smoking status via questionnaire. Current smokers included participants who had: (1) smoked at least 100 cigarettes in their lifetime and reported smoking at the time of the interview, or (2) quit at diagnosis and within 1 year prior to diagnosis. Non-current smokers included former smokers (smoked at least 100 cigarettes in lifetime and who quit at least 1 year prior to diagnosis) and never smokers; 62.7% of non-smokers were never smokers. In BRFSS, a computed smoking status was used to assess current smoking status, originally with 4 levels: everyday smoker, someday smoker, former and non-smoker, recoded to “current (everyday and someday) vs. not current (former and non-smoker)”. For anthropometry, CBCS3 BMI was based on nurse measured anthropometric data and was measured in weight in kilograms/height in meters squared. In BRFSS, BMI was calculated and categorized as: neither overweight nor obese, overweight and obese. The variable was derived from self-report height and weight. The final variable for BMI was: less than or equal to 25, 25 to 30, and greater than 30.

### Access to medical care factors: CBCS3

Access to medical care factors included currently insured (yes vs. no), insurance type, rural residence (yes vs. no), financial issues (yes vs. no), transportation issues (yes vs. no), and job loss (yes vs. no). Participants were asked about insurance, insurance type, rural status, and job loss at baseline. Private insurance was defined as private health insurance from employer or workplace or that of the participant’s husband or partner. Other types included Medicaid, Medicare, other and none. For job loss, participants were asked if they had lost their job to diagnosis of breast cancer. Participants were also asked at approximately 18 months follow-up telephone survey from baseline if they could not see a doctor because of financial and transportation issues.

### Tumor characteristics: CBCS3

Tumor size ( $\leq 2$  cm vs.  $> 2$  cm), nodal status (positive vs. negative) and histologic grade (I and II vs. III) were obtained from medical records abstraction. Estrogen receptor (ER) status, progesterone receptor (PR) status, human epidermal growth factor receptor 2 (HER2), and Triple-Negative (TN) status (positive vs. negative) were obtained from pathology reports abstraction for 98% of cases in CBCS3. IHC staining was done for the remaining 2% of cases without medical record data [10]. HER2 was derived from Immunohistochemical (IHC) and/or Fluorescence In Situ Hybridization (FISH) assay. A positive ER or PR was defined as  $> 10\%$  cut point. Borderlines, which represented 8% of cases, were included with negative status consistent with the recommendation of Allott et al., which showed that ER borderlines were more likely to have non-luminal subtype [11].

### Statistical analysis

Weighted percentages were calculated for selected participants’ SES and comorbidity factors by race and age in CBCS3 and NC BRFSS. We examined SES, comorbidities, access to medical care and the individual tumor characteristic factors using latent class analysis (LCA) to identify groups of individuals based on numerical factors using PROC LCA a SAS package [12]. LCA identifies unobservable, or latent, subgroups within a population that are mutually exclusive and exhaustive [13]. The model probabilistically groups each observation into a latent class, a variable indicating underlying subgroups of individuals based on observed characteristics. An iterative approach to parameter estimation using expectation–maximization (EM) for maximum-likelihood (ML) estimation generated estimates of all model parameters and item-response probabilities of class assignment. We used several criteria to determine the number of classes from the ML solution using 100,000 sets of starting values. We examined the  $G^2$  likelihood-ratio test statistic, Akaike’s information criterion (AIC) and Bayesian Information Criterion (BIC), a goodness-of-fit measure to find more parsimonious models, where a smaller AIC and BIC suggest a better model fit. Additionally, we evaluated entropy, where higher values reflect better classification. We then expanded model specification for multiple-groups LCA to examine differences in risk factor patterns by race and age. Here, we tested for measurement invariance across groups for differences between younger and older and black and white women. We compared a series of latent class models to determine the optimal model for parsimony and model fit using the criteria described above.

We examined the distribution of latent class categories stratified by race and age. We used logistic regressions and polytomous logistic regressions to estimate odd ratios

(ORs) and 95% confidence intervals (CIs) as the measure of association between SES/comorbidities, access to care patterns, individual tumor factors latent classes and race/age. All statistical analyses were conducted with SAS statistical package version 9.3 (SAS Institute, Inc., Cary, NC). *P* values were for a two-sided test with an alpha of 0.05 for statistical significance.

## Results

### CBCS3 in context of the state of North Carolina

The CBCS3 was designed to assess race and age differences in breast cancer outcomes, and therefore, oversampled black and younger women. However, our findings show that after weighting for randomized recruitment, CBCS3 shows generally similar distributions of socioeconomic and comorbidity variables as the BRFSS data for the state as a whole (Fig. 1). Racial and age-specific patterns for marital status, diabetes, heart disease, smoking status and income were very similar in both datasets. Compared to white women, black women were less likely to be married, more commonly resided in non-rural areas, were more likely to be obese, had lower income and had lower proportions of health insurance and education. Compared to older women, younger women were more likely to earn higher levels of income and have higher level of education. Some expected differences were observed for rurality, given that the CBCS3 sampled only from the most populous 44-counties in the state. The CBCS3 population also had a slightly greater proportion of women with current health insurance and post high school education. White women in CBCS3 were somewhat more likely to be in the highest category of income compared to white women in BRFSS. In both populations, each of the 9 categorical variables for SES and comorbidity had at least 10% prevalence, except heart disease shown in Supplement Tables 1A and 1B.

Compared to white women in CBCS3, black women in CBCS had lower proportions of private health insurance and rural residence, but higher proportions of Medicare and both financial and transportation barriers to care (Supplemental Table 2). Compared to older white women, younger white women and both younger and older black women in CBCS had higher proportions of Medicaid, job loss due to breast cancer diagnosis, and financial and transportation barriers to care.

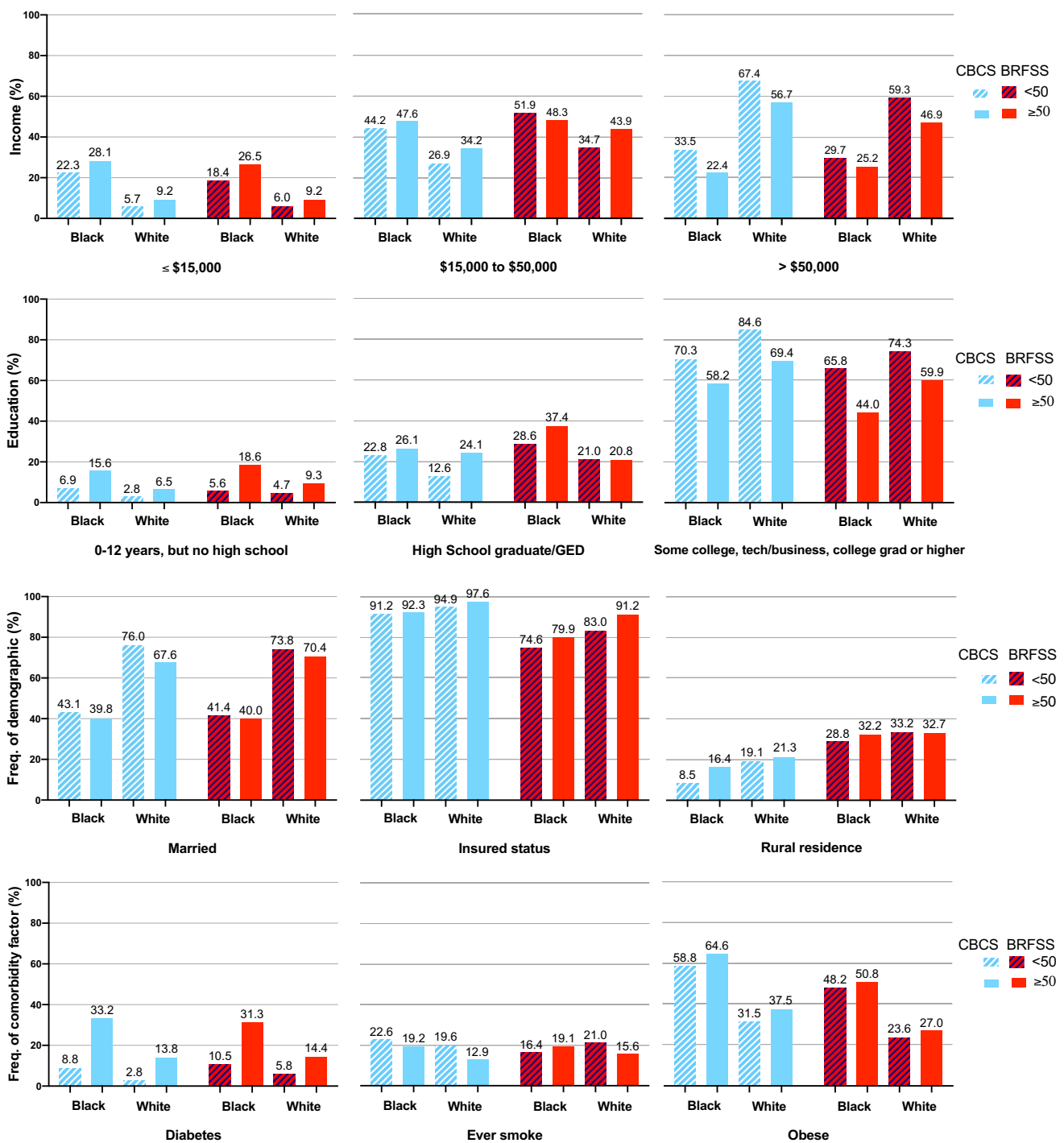
Barriers to cancer care and clinical characteristics of tumors are not included in BRFSS, but we assessed distributions of these variables in CBCS3. Supplemental Table 2 shows individual barriers to care by race and age and Supplemental Table 3 shows clinical characteristics. Black younger women had more financial and transportation issues

and experienced more job loss due to breast cancer diagnosis. Overall, older white women had the most favorable tumor characteristics, with more aggressive features seen for both younger and black women (Supplemental Fig. 1).

### Latent classes and patterns by race and age

Given the multiplicity of variables that comprise socioeconomic status, barriers to care, and tumor biology, respectively, and the tendency for characteristics to aggregate within individuals, we sought to develop latent classes based on these variables. We a priori identified three separate latent class domains: (1) SES/comorbidity factors, (2) barrier to care factors and (3) tumor characteristics. We chose these domains because functionally they relate separately to the lived experience and exposure, access factors, which we considered to be most intervenable, and clinical intervention factors. For the SES/comorbidity factors, a model with two latent classes had lowest AIC, BIC and likelihood-ratio  $G^2$  statistic compared to one, three, four, five or six-class models. For the barriers to care, two classes were optimal, and for tumor characteristics, four latent classes had the lowest AIC, BIC and likelihood-ratio  $G^2$  statistic. We examined contingency tables (Supplemental Tables 4–6) of the latent class posterior probabilities. We observed no correlation among the latent class groups and no further collapsibility, suggesting that these three latent classes were a stable classification solution.

Social, comorbidity, access to care and tumor factors determined distinct class membership. Figure 2 shows the relative frequency of each latent class and the posterior probabilities that members of each latent class endorsed each factor category in the models for: SES and comorbidity factors (panel A), barriers to care (panel B) and tumor characteristics (panel C). For the two-class model measuring SES and comorbidity factors, 54% were in the *High SES/Low Comorbidity* latent class, which was characterized by a high probability of the highest categories of income, education and being married and low probability of comorbidities. For the two-class model for barriers to care, 87% of women were in the *Less Barriers* latent class, with lower probability of being uninsured, having financial or transportation issues, residing in rural areas, or job loss. For the four-class model measuring tumor characteristics, 40% of the women were in the *HR+/HER2-/Node Negative* latent class, which was characterized as HR+, HER2–, smaller tumor size, node negative and lower grade tumors. The *HER2+/Higher Grade* class (10%) was characterized by highest probabilities of having HER2+ breast cancers and higher probabilities of higher grade, node positive and larger tumor size. The *HR+/HER2-/Larger Tumor/Node Positive* class (29%) was characterized by highest probabilities of having HR+, larger tumor size and node positive tumors. Lastly, the *TNBC/*



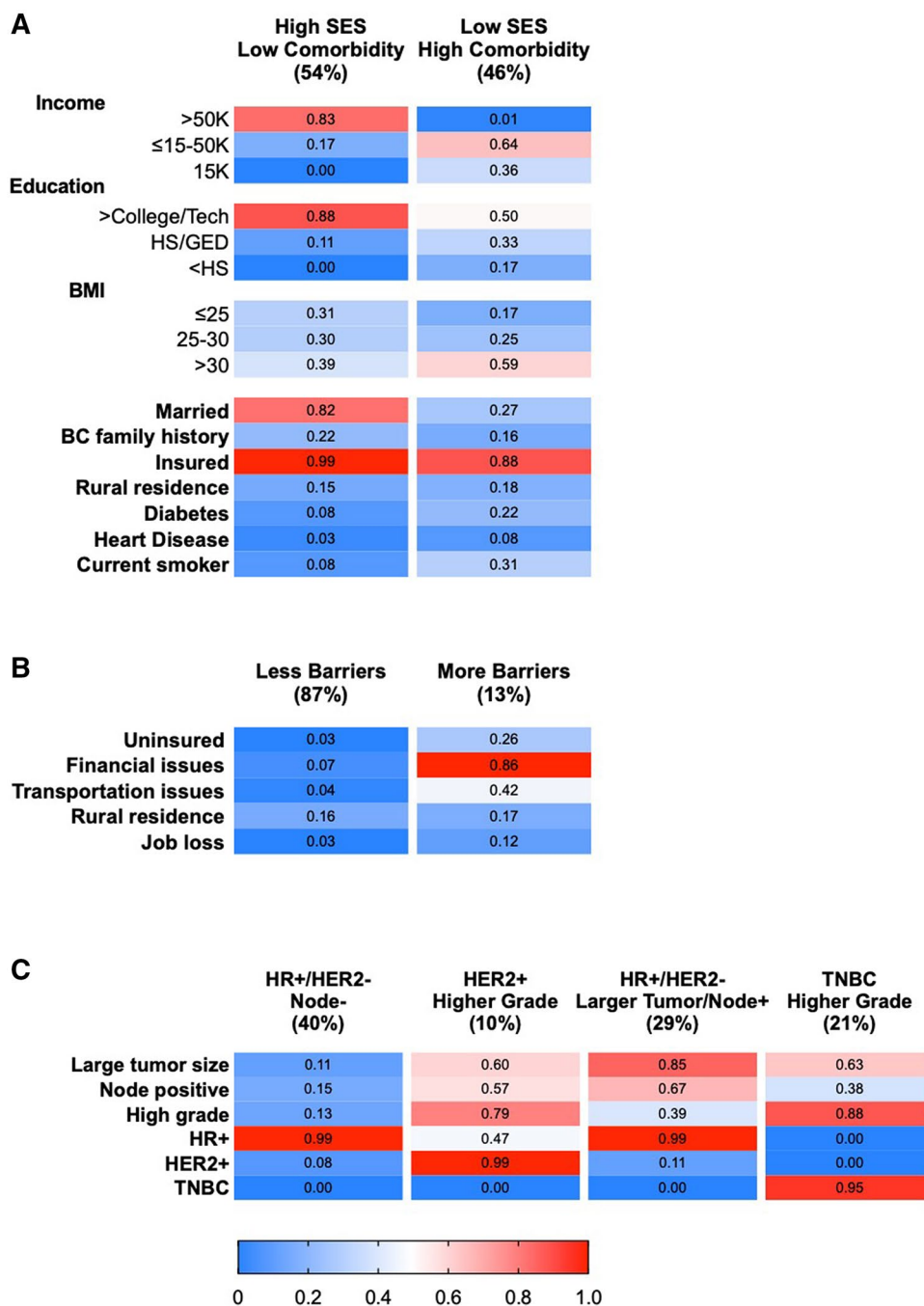
**Fig. 1** Distributions of categorical demographic and comorbid characteristics for women in the Carolina Breast Cancer Study 3 (2008–2013) and Behavioral Risk Factor Surveillance System (2010), by age and race

*Higher Grade* class (21%) was characterized by highest probabilities of TNBC and higher grade tumors.

Race and age were associated with latent class membership for all three domains we evaluated. As shown in Fig. 3, when controlling for the effect of age, black race was positively associated with low SES (OR 6.3,

95% CI 5.2, 7.8; vs. high SES), more barriers to medical care (OR 5.6, 95% CI 3.9, 8.1; vs. less barriers) and more aggressive aggregated tumor characteristics (e.g., TNBC OR 3.8, 95% CI 3.0, 4.8) compared to white women. Controlling for race, younger age was negatively associated

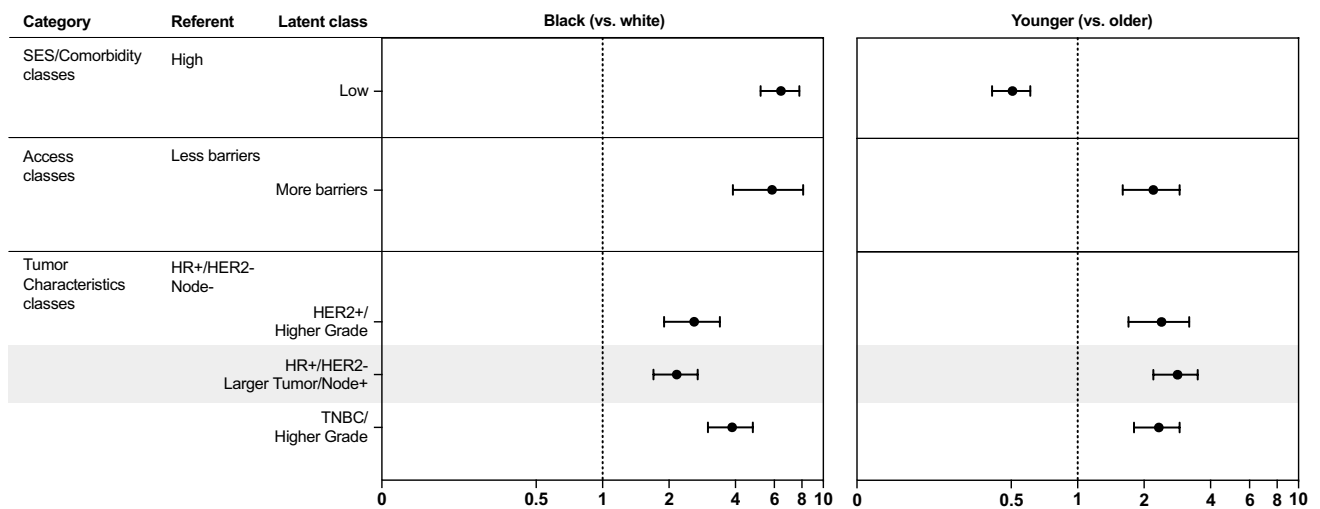
**Fig. 2** Probabilities of responses to items for **a** two-class model measuring SES/comorbidities, **b** two-class model measuring access to medical care **c** four-class model measuring tumor characteristics



with low SES (OR 0.5, 95% CI 0.41, 0.61; vs. high SES), but positively associated with more barriers to medical care (OR 2.1, 95% CI 1.6, 2.9; vs. less barriers) and more aggressive aggregated tumor characteristics (e.g., TNBC OR 2.3, 95% CI 1.8, 2.9). Further cross-classification on race and age also showed strong determinants of class membership patterns (Table 1). Taken together, these models suggest a compounding of high risk SES, access, and tumor biological factors for younger and black women.

### Discussion

The CBCS3 population-based cohort of women with breast cancer, which oversampled younger and black women, has helped further elucidate the complexity of race and age disparities. Women in CBCS3 were representative of women as a whole in North Carolina, with women in CBCS3 and BRFSS showing similar distributions of SES and comorbidity factors. Slightly higher income, education



**Fig. 3** Odds ratios and 95% confidence intervals for association of race and age with specific latent classes, Carolina Breast Cancer Study 3 (2008–2013)

**Table 1** Distribution and odds ratios with 95% confidence intervals of latent classes, Carolina Breast Cancer Study 3 (2008–2013) by age and race

	All cases (n = 2998)		Race and age (years)							
			White				Black			
			< 50 (n = 751)		≥ 50 (n = 752)		< 50 (n = 741)		≥ 50 (n = 754)	
	No	(%)	No	(%)	No	(%)	No	(%)	No	(%)
<b>SES/Comorbidity latent classes</b>										
Low SES/high comorbidity	1,428	(47.6)	134	(9.4)	259	(18.1)	455	(31.9)	580	(40.6)
High SES/low comorbidity	1,570	(52.4)	617	(39.3)	493	(31.4)	286	(18.2)	174	(11.1)
Low vs. high, OR (95% CI)			<b>0.5</b>	<b>(0.4–0.6)</b>	1.0	Ref	<b>3.2</b>	<b>(2.5–4.1)</b>	<b>6.2</b>	<b>(4.7–8.1)</b>
<b>Access to care latent classes</b>										
Less barriers to care	2,514	(84.1)	709	(28.2)	737	(29.3)	482	(19.2)	574	(22.8)
More barriers to care	476	(15.9)	42	(8.8)	15	(3.2)	258	(54.2)	180	(37.8)
Less vs. more, OR (95% CI)			<b>3.3</b>	<b>(1.6–6.4)</b>	1.0	Ref	<b>15.3</b>	<b>(7.8–29.9)</b>	<b>7.8</b>	<b>(4.1–15.4)</b>
<b>Tumor characteristics latent classes<sup>a</sup></b>										
HR+/HER2–/node negative	1,091	(36.5)	290	(26.6)	419	(38.4)	117	(10.7)	265	(24.3)
HER2+/higher grade	335	(11.2)	86	(25.7)	59	(17.6)	99	(29.5)	91	(27.2)
HR+/HER2–/larger tumor/node positive	944	(31.6)	259	(27.4)	181	(19.2)	301	(31.9)	203	(21.5)
TNBC/higher grade	617	(20.7)	115	(18.6)	90	(14.6)	222	(36.0)	190	(30.8)
HER2+ vs. HR+/HER2–/Node–, OR (95% CI)			<b>2.6</b>	<b>(1.9–3.6)</b>	1.0	Ref	<b>6.3</b>	<b>(4.3–9.1)</b>	<b>2.0</b>	<b>(1.4–2.8)</b>
HR+ vs. HR+/HER2–/Node–, OR (95% CI)			<b>2.3</b>	<b>(1.5–3.6)</b>	1.0	Ref	<b>6.3</b>	<b>(3.9–10.0)</b>	<b>2.6</b>	<b>(2.5–3.8)</b>
TNBC vs. HR+/HER2–/Node–, OR (95% CI)			<b>2.0</b>	<b>(1.5–2.9)</b>	1.0	Ref	<b>9.0</b>	<b>(6.1–13.0)</b>	<b>3.4</b>	<b>(2.5–4.6)</b>

Bold indicates the significance of 95% confidence intervals

Latent class domains not mutually adjusting for one another

SES Socioeconomic status, HER2+ human epidermal growth factor receptor 2 positive, TNBC triple-negative breast cancer

<sup>a</sup>N = 2,987

and non-rural residence in CBCS3 compared to BRFSS as a whole likely reflects CBCS3's sampling of more populous counties. It may also reflect differences in participation demand for the two studies, with CBCS3 requiring a home visit, an extensive follow-up schedule and BRFSS including just a single telephone interview. Many individual variables differed by race and age in the CBCS3 and to evaluate person-centered patterns, we used latent class analysis to show that black women were much more likely to have low SES, more aggressive aggregated tumor characteristics, and experience more barriers to medical care. Compared to older women, younger women were more likely to have higher SES, but had more aggressive aggregated tumor characteristics and experienced more barriers to medical care. Constructing and examining these covariates in aggregate helps to summarize a deeper understanding and a more nuanced view of race and age in breast cancer disparities.

Based on the previous literature, racial differences observed in CBCS3 portend important differences in mortality. Racial differences in diagnosis, treatment and mortality have been shown to be greater among women who are unmarried [14], live in more rural areas [15], have lower SES [16–19], and have more comorbid conditions [20]. Our observation of higher obesity rates in CBCS3 patients compared to the NC BRFSS population may be explained by obesity's status as a risk factor for cancer [21–23]. An additional consideration is the difference in the measurement of height and weight in both studies. A nurse in CBCS3 objectively measured height and weight, whereas in BRFSS, these measures were self-reported, which tends to result in underreporting of obesity [24, 25]. Social, economic and cultural barriers to care, including insurance and cancer-care cost, have also been shown to contribute to racial disparities [26–28]. Our results are consistent with the well-established literature on racial differences in tumor biology. Black women are at much greater risk of higher grade, ER-negative and triple-negative breast cancers [29–31], even after controlling for age and stage. Younger and black women are more likely to have subtypes of breast cancer tumors that are more aggressive, treatment resistant and without clear molecular targets for treatment [32]. A published review highlights how both tumor biology and quality of care potentially increase the racial mortality gap [33]. Yet, many manuscripts continue to consider tumor biology and access separately, with little integration. Much remains poorly understood about the cancer experience of younger black women and thus, the disparity is widening [4]. It is important to develop integrated approaches to studying breast cancer mortality.

Breast cancer mortality is higher in poorer communities regardless of race/ethnicity [34]. Previous analyses of racial identity and socioeconomic and social environment factors

have made it evident that race is a sociopolitical construct in addition to a biological or genetic characteristic [35]. Socioeconomic disadvantages are more prevalent in Black communities compared to white communities, including insurance, income, education, economic, environmental and structural factors [36]. Prevalences for modifiable factors, such as cigarette smoking and obesity, are higher with less education and below poverty level among women [34], in part, because targeted cigarette company advertisements [37], environmental changes and decreases in physical activity [38].

Previous population-based breast cancer studies have used area-level SES composite scores (e.g., area-level education or income and/or individual insurance) [28, 39, 40], but such approaches have varying validity by dataset because SES manifests in different ways in distinct populations [41]. Area-level SES has been used as a continuous index of social and economic indicators, but it is difficult to interpret because it lacks an identifiable unit and the assumption of a constant linear relationship may not be valid. Additionally, SES measured at the area-level does not fully capture individual variation, which is needed to inform intervention approaches. Palumbo et al. using, latent class models, identified factors of advantage and disadvantage neighborhood SES and measured the association of neighborhood SES, race and breast cancer tumor characteristics. Consistent with our findings, they found that high disadvantage neighborhood SES (i.e., higher proportions of neighborhoods with single with dependents, below poverty line, low vehicle access, black race) was associated with more aggressive tumor characteristics (i.e., lower proportions of Luminal A breast cancers, earlier stage, smaller size and lower grade) [42]. Importantly, this study also found lack of concordance when comparing the latent class variables with a continuous neighborhood SES index, suggesting that SES characteristics are better represented by multiple latent classes than by a single index. Indicators of SES are not perfectly correlated, which makes LCA an advantageous alternative measurement tool in assessing multidimensional SES factors in cancer outcomes research. Strong links between SES, tumor characteristics, and access patterns suggest that analyses that integrate biological and access characteristics are important for understanding mortality disparities. Many cancer biology studies lack data on SES and access variables, and few studies have adequately integrated cancer biology and barriers to care in understanding disparities. The current analysis extends previous studies that by incorporating reported barriers to care, tumor characteristics patterns, individual level SES, and by examining these differences by race, with race as a covariate not as a characteristic of the latent class models, and by race and age simultaneously.

Our study should be interpreted with some limitations in mind. First, we were unable to restrict our BRFSS



analysis to the same sampling region used for CBCS3 because we did not have access to the BRFSS geocoded data so some comparisons may reflect regional differences in the state rather than differences between the study population and the source population. Second, it is possible that we are underestimating some access and biological disparities relative to the most underserved patients in the state. Notably, over 90% of CBCS3 women have some health insurance, higher than reported rates nationally and in NC during this time period. However, we did include many uninsured patients and because our goal was to identify patterns in overall SES and comorbidity factors, and we do not expect that the latent classes would be substantially different when expanding to the entirety of NC. Third, the access to care measures, including financial issues, are self-reported. As such, they may represent heterogeneous types of financial issues and varying sensitivity to financial stressors. Fourth, we also considered only a limited number of potential domains that could be associated with mortality disparities, however, we were able to decrease dimensionality of these domains considerably. Our study has several key strengths including the population-based study design of CBCS3, which oversampled young and black women with breast cancer. Another strength is the depth of data on socioeconomic, demographic, access and tumor biology. Finally, our population size was substantial enough to be able to find stable latent classes.

## Conclusion

Overall, previous studies have shown that breast cancer mortality is a complex combination of race, SES, and genetic factors. This paper is a step towards more fully capturing and comprehensively examining the interaction and patterning of all these important indicators that reflect health disparity.

Probabilistic model-based clustering methods identified unique patterns based on SES, barriers to care, and biological factors. Together, these factors can help specify interventions to reduce breast cancer mortality disparities. Future research should evaluate how these factors relate to outcomes, such as treatment delays and recurrence.

**Acknowledgments** This work was supported by a Grant from the UNC Lineberger Comprehensive Cancer Center funded by the University Cancer Research Fund (LCCC2017T204), Susan G. Komen Graduate Training in Disparities Research, the National Cancer Institute of the National Institutes of Health (P50-CA58223, U01-CA179715, T32-CA057726), and the National Institute of Environmental Health Sciences of the National Institutes of Health (P30-ES010126).

## Compliance with ethical standards

**Conflicts of interest** The authors declare that they have no conflict of interest.

## References

1. DeSantis CE, Fedewa SA, Sauer AG, Kramer JL, Smith RA, Jemal A (2016) Breast cancer statistics, 2015: convergence of incidence rates between black and white women. *CA Cancer J Clin* 66(1):31–42. <https://doi.org/10.3322/caac.21320>
2. DeSantis CE, Siegel RL, Sauer AG, Miller KD, Fedewa SA, Alcaraz KI, Jemal A (2016) Cancer statistics for African Americans, 2016: progress and opportunities in reducing racial disparities. *CA Cancer J Clin* 66(4):290–308. <https://doi.org/10.3322/caac.21340>
3. National Center for Health Statistics, National Vital Statistics System Table 26 (2017) Death rates for malignant neoplasm of breast among females, by race, Hispanic origin and age: United States 2017. US Department of Health and Human Services, Centers for Disease Control and Prevention, Hyattsville, MD
4. Polyak K (2011) Heterogeneity in breast cancer. *J Clin Invest* 121(10):3786–3788. <https://doi.org/10.1172/jci60534>
5. Dunn BK, Agurs-Collins T, Browne D, Lubet R, Johnson KA (2010) Health disparities in breast cancer: biology meets socioeconomic status. *Breast Cancer Res Treat* 121(2):281–292. <https://doi.org/10.1007/s10549-010-0827-x>
6. Bergman LR, Mabnsson D, El-Khoury BM (2003) Studying individual development in an interindividual context: a person-oriented approach. Lawrence Erlbaum Associates, Mahwah, NJ
7. Rosner B, Colditz GA, Willett WC (1994) Reproductive risk factors in a prospective study of breast cancer: the Nurses' Health Study. *Am J Epidemiol* 139(8):819–835
8. Andaya AA, Enewold L, Horner MJ, Jatoi I, Shriver CD, Zhu K (2012) Socioeconomic disparities and breast cancer hormone receptor status. *Cancer Causes Control* 23(6):951–958. <https://doi.org/10.1007/s10552-012-9966-1>
9. Warner ET, Colditz GA, Palmer JR, Partridge AH, Rosner BA, Tamimi RM (2013) Reproductive factors and risk of premenopausal breast cancer by age at diagnosis: are there differences before and after age 40? *Breast Cancer Res Treat* 142(1):165–175. <https://doi.org/10.1007/s10549-013-2721-9>
10. Allott EH, Cohen SM, Geradts J, Sun X, Khoury T, Bshara W, Zirpoli GR, Miller CR, Hwang H, Thorne LB, O'Connor S, Tse CK, Bell MB, Hu Z, Li Y, Kirk EL, Bethea TN, Perou CM, Palmer JR, Ambrosone CB, Olshan AF, Troester MA (2016) Performance of three-biomarker immunohistochemistry for Intrinsic Breast Cancer Subtyping in the AMBER Consortium. *Cancer Epidemiol Biomarkers Prev* 25(3):470–478. <https://doi.org/10.1158/1055-9965.epi-15-0874>
11. Allott EH, Geradts J, Cohen SM, Khoury T, Zirpoli GR, Bshara W, Davis W, Omilian A, Nair P, Ondracek RP, Cheng TD, Miller CR, Hwang H, Thorne LB, O'Connor S, Bethea TN, Bell ME, Hu Z, Li Y, Kirk EL, Sun X, Ruiz-Narvaez EA, Perou CM, Palmer JR, Olshan AF, Ambrosone CB, Troester MA (2018) Frequency of breast cancer subtypes among African American women in the AMBER consortium. *Breast Cancer Res* 20(1):12. <https://doi.org/10.1186/s13058-018-0939-5>
12. PROC LCA, PROC LTA (Version 1.3.2) [Software] (2015) The Methodology Center, Penn State, University Park. <https://methodology.psu.edu/>.

13. Lanza ST, Collins LM, Lemmon DR, Schafer JL (2007) PROC LCA: a SAS procedure for latent class analysis. *Struct Equ Model Multidiscip J* 14(4):671–694
14. Hsu CD, Wang X, Habif DV Jr, Ma CX, Johnson KJ (2017) Breast cancer stage variation and survival in association with insurance status and sociodemographic factors in US women 18 to 64 years old. *Cancer* 123(16):3125–3131. <https://doi.org/10.1002/cncr.30722>
15. Mobley LR, Kuo TM, Clayton LJ, Evans WD (2009) Mammography facilities are accessible, so why is utilization so low? *Cancer Causes Control* 20(6):1017–1028. <https://doi.org/10.1007/s10552-009-9295-1>
16. Bigby J, Holmes MD (2005) Disparities across the breast cancer continuum. *Cancer Causes Control* 16(1):35–44. <https://doi.org/10.1007/s10552-004-1263-1>
17. Gerend MA, Pai M (2008) Social determinants of Black-White disparities in breast cancer mortality: a review. *Cancer Epidemiol Biomarkers Prev* 17(11):2913–2923. <https://doi.org/10.1158/1055-9965.epi-07-0633>
18. Parise CA, Caggiano V (2013) Disparities in race/ethnicity and socioeconomic status: risk of mortality of breast cancer patients in the California Cancer Registry, 2000–2010. *BMC Cancer* 13:449. <https://doi.org/10.1186/1471-2407-13-449>
19. Bradley CJ, Given CW, Roberts C (2002) Race, socioeconomic status, and breast cancer treatment and survival. *J Natl Cancer Inst* 94(7):490–496
20. Tammemagi CM (2007) Racial/ethnic disparities in breast and gynecologic cancer treatment and outcomes. *Curr Opin Obstet Gynecol* 19(1):31–36. <https://doi.org/10.1097/GCO.0b013e3280117cf8>
21. Renehan AG, Tyson M, Egger M, Heller RF, Zwahlen M (2008) Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet* 371(9612):569–578. [https://doi.org/10.1016/s0140-6736\(08\)60269-x](https://doi.org/10.1016/s0140-6736(08)60269-x)
22. Munsell MF, Sprague BL, Berry DA, Chisholm G, Trentham-Dietz A (2014) Body mass index and breast cancer risk according to postmenopausal estrogen-progestin use and hormone receptor status. *Epidemiol Rev* 36:114–136. <https://doi.org/10.1093/epirev/mxt010>
23. Flegal KM, Kruszon-Moran D, Carroll MD, Fryar CD, Ogden CL (2016) Trends in obesity among adults in the United States, 2005 to 2014. *JAMA* 315(21):2284–2291. <https://doi.org/10.1001/jama.2016.6458>
24. Connor Gorber S, Tremblay M, Moher D, Gorber B (2007) A comparison of direct vs. self-report measures for assessing height, weight and body mass index: a systematic review. *Obesity Rev* 8(4):307–326. <https://doi.org/10.1111/j.1467-789X.2007.00347.x>
25. Nawaz H, Chan W, Abdulrahman M, Larson D, Katz DL (2001) Self-reported weight and height: implications for obesity research. *Am J Prev Med* 20(4):294–298
26. Ashing-Giwa KT, Padilla G, Tejero J, Kraemer J, Wright K, Coscarelli A, Clayton S, Williams I, Hills D (2004) Understanding the breast cancer experience of women: a qualitative study of African American, Asian American, Latina and Caucasian cancer survivors. *Psycho-oncology* 13(6):408–428. <https://doi.org/10.1002/pon.750>
27. Wheeler SB, Spencer JC, Pinheiro LC, Carey LA, Olshan AF, Reeder-Hayes KE (2018) Financial impact of breast cancer in black versus white women. *J Clin Oncol* 36(17):1695–1701. <https://doi.org/10.1200/jco.2017.77.6310>
28. DeSantis C, Jemal A, Ward E (2010) Disparities in breast cancer prognostic factors by race, insurance status, and education. *Cancer Causes Control* 21(9):1445–1450. <https://doi.org/10.1007/s10552-010-9572-z>
29. Anderson WF, Chatterjee N, Ershler WB, Brawley OW (2002) Estrogen receptor breast cancer phenotypes in the surveillance, epidemiology, and end results database. *Breast Cancer Res Treat* 76(1):27–36
30. Smith-Bindman R, Miglioretti DL, Lurie N, Abraham L, Barbash RB, Strzelczyk J, Dignan M, Barlow WE, Beasley CM, Kerlikowske K (2006) Does utilization of screening mammography explain racial and ethnic differences in breast cancer? *Ann Intern Med* 144(8):541–553
31. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, Karaca G, Troester MA, Tse CK, Edmiston S, Deming SL, Geradts J, Cheang MCU, Nielsen TO, Moorman PG, Earp HS, Millikan RC (2006) Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA* 295(21):2492–2502. <https://doi.org/10.1001/jama.295.21.2492>
32. Dunnwald LK, Rossing MA, Li CI (2007) Hormone receptor status, tumor characteristics, and prognosis: a prospective cohort of breast cancer patients. *Breast Cancer Res* 9(1):R6. <https://doi.org/10.1186/bcr1639>
33. Daly B, Olopade OI (2015) A perfect storm: how tumor biology, genomics, and health care delivery patterns collide to create a racial survival disparity in breast cancer and proposed interventions for change. *CA Cancer J Clin* 65(3):221–238. <https://doi.org/10.3322/caac.21271>
34. Ward E, Jemal A, Cokkinides V, Singh GK, Cardinez C, Ghafoor A, Thun M (2004) Cancer disparities by race/ethnicity and socioeconomic status. *CA Cancer J Clin* 54(2):78–93
35. Williams DR, Mohammed SA, Shields AE (2016) Understanding and effectively addressing breast cancer in African American women: unpacking the social context. *Cancer* 122(14):2138–2149. <https://doi.org/10.1002/cncr.29935>
36. Williams DR, Mohammed SA, Leavell J, Collins C (2010) Race, socioeconomic status, and health: complexities, ongoing challenges, and research opportunities. *Ann N Y Acad Sci* 1186:69–101. <https://doi.org/10.1111/j.1749-6632.2009.05339.x>
37. Tobacco use among US Racial/Ethnic Minority Groups—African-Americans, American Indians, and Alaska Natives, Asian-Americans and Pacific Islanders, and Hispanics: a report of the surgeon general (1998)
38. Hill JO, Wyatt HR, Reed GW, Peters JC (2003) Obesity and the environment: where do we go from here? *Science* 299(5608):853–855. <https://doi.org/10.1126/science.1079857>
39. Akinyemiju TF, Pisu M, Waterbor JW, Altekruse SF (2015) Socioeconomic status and incidence of breast cancer by hormone receptor subtype. *Springerplus* 4:508. <https://doi.org/10.1186/s40064-015-1282-2>
40. Sineshaw HM, Gaudet M, Ward EM, Flanders WD, Desantis C, Lin CC, Jemal A (2014) Association of race/ethnicity, socioeconomic status, and breast cancer subtypes in the National Cancer Data Base (2010–2011). *Breast Cancer Res Treat* 145(3):753–763. <https://doi.org/10.1007/s10549-014-2976-9>
41. Messer LC, Laraia BA, Kaufman JS, Eyster J, Holzman C, Culhane J, Elo I, Burke JG, O'Campo P (2006) The development of a standardized neighborhood deprivation index. *J Urban Health* 83(6):1041–1062. <https://doi.org/10.1007/s11524-006-9094-x>
42. Palumbo A, Michael Y, Hyslop T (2016) Latent class model characterization of neighborhood socioeconomic status. *Cancer Causes Control* 27(3):445–452. <https://doi.org/10.1007/s10552-015-0711-4>

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